ISBN: 978-1-7329562-6-1

Estd. in 2001

Conferred Autonomous Status by University Grants Commission(UGC) for 10 years w.e.f. A.Y. 2019-20

- ISO 9001:2015 Certified Institute
- NBA Accredited Programs
- NAAC Accredited Institute with 'A' Grade
- AICTE-CII Survey rating in Platinum category for Industry linkages
- Amongst Top 200 Colleges in the Country Ranked 193rd in NIRF India Ranking 2019 in Engineering Category
- 68th & 78th in All India Rank by Outlook survey published in June 2019 & May 2018 respectively

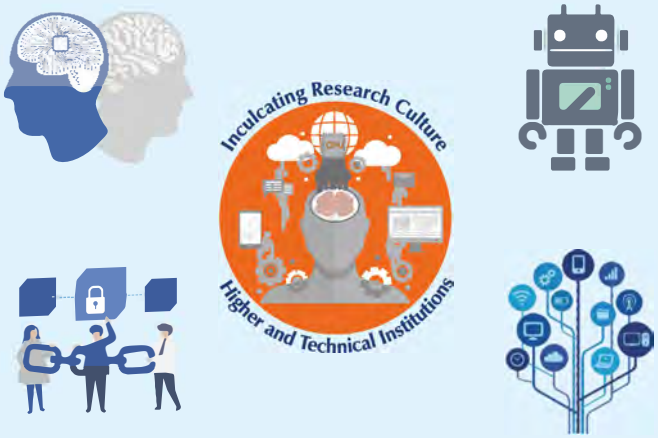
MCW 2020 Invention & Innovation

Proceedings of

International Conference on Intelligent Computing and Networking - 2020

IC-ICN - 2020

Organized by: **Department of Computer Engineering & Information Technology**



Editors

Dr. Sheetal Rathi

Dr. Rajesh Bansode

Mr. Vikas Kaul

Mrs. Shiwani Gupta

Dr. Anand Khandare

Mr. Aaditya Desai

In Association With

Springer

Jagdu Singh Charitable Trust's (Regd.)

THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY

Autonomous College Affiliated to University of Mumbai

Approved by All India Council for Technical Education(AICTE) and Government of Maharashtra

A - Block, Thakur Educational Campus, Shyamnarayan Thakur Marg,
Thakur Village, Kandivali (East), Mumbai - 400 101

Tel.: 022-6730 8000 / 8106 / 8107 Telefax: 022-2846 1890 • Email: tcet@thakureducation.org

• Website: www.tcetmumbai.in www.thakureducation.org

Estd. in 2001

Conferred Autonomous Status by University Grants Commission(UGC) for 10 years w.e.f. A.Y. 2019-20

ISO 9001:2015 Certified Institute | NBA Accredited Programs | NAAC Accredited Institute with 'A' Grade | AICTE-CII Survey rating in Platinum category for Industry linkages | Amongst Top 200 Colleges in the Country Ranked 193rd in NIRF India Ranking 2019 in Engineering Category | 68th & 78th in All India Rank by Outlook survey published in June 2019 & May 2018 respectively

11th CONFERENCE

IC-ICN – 2020 Conference Proceeding

MCW 2020 Invention & Innovation

MULTICON-W 2020

A platform for Multiple Conferences and Workshops

11th International & National Conferences and Workshops

In Association with

Springer

International Conference on Intelligent Computing and Networking – 2020

Chief Patron
Mr. V. K. Singh, Chairman, TEG

Patrons
Mr. Jitendra R. Singh, Trustee, TEG
Mr. Karan V. Singh, CEO, TCET, TIMSR, TIMSCDR

Program Chair
Dr. B. K. Mishra, Principal, TCET

Program Co- Chair
Dr. Deven Shah, Vice-Principal

Advisory Committee

1 Dr. Dusko Lukac CCEC – Centre of Competence, Germany.
2 Dr. Alveen Singh, Durban University of Technology, South Africa.
3 Prof. Valentina Emilia Balas, Aurel Vlaicu University of Arad.
4 Dr. Selwyn Piramuthu Professor, University of Florida.
5 Preggy Ready, LMS Administrator, Durban University of Technology.
6 Prof. Bharat Bhargava Purdue University Department of Computer Sciences, Indiana.
7 Dr. Omer F. Rana, Professor, Cardiff University, UK.
8 Dr. Javed I. Khan, Professor & Chair, Department of Computer Science, Kent State University, Kent.
9 Prof. Harpreet Singh, College of Engineering, Wayne State University Detroit, MI.
10 Dr. Arvind Vasu Professor, Clarion University, Clarion Pennsylvania, United States.
11 Ferdinand-James D Professor, University of the West Indies, Trinidad and Tobago, West Indies.
12 Mr. Gautam Mishra Software Engineer, IBM, USA.
13 Mr. Sachine Armal Software Engineer, GE Digital, USA.
14 Mr. Nitin Hiwrale Software Engineer, USA.
15 Dr. Sateesh Kumar, Associate Professor, IIT, Roorkee.
16 Dr. Padmaja Joshi, Director CDAC Mumbai, India.
17 Dr. Suneeta Agarwal, Professor, NIT, Allahabad.
18 Dr. Mustaq Ahmed , Associate Professor , NIT Jaipur.
19 Dr. Sambit Kumar Mishra ,Professor, Department of Computer Sc.&Engg. Head (R&D), Gandhi Institute for Education and Technology, Baniatangi, Khordha.
20 Dr. Vijay Bhaskar Semwal, Assistant Professor, NIT, Bhopal, MP.

Technical Chair
Dr. R.R. Sedamkar,Dean Academic
Dr. KamalShah,Dean R&D

Convener
Dr. SheetalRathi, HOD-COMP

Joint Convenor
Dr. Rajesh Bansode,HOD-IT

Co-convenors
Mrs. Shiwani Gupta,Deputy HOD-COMP
Mr. Vikas Kaul, Deputy HOD-IT

Overall Coordinator
Dr. Anand Khandare - Associate Professor – COMP
Mr. Aditya Desai – Assistant Professor - IT

Publication Committee
Dr. Sheetal Rathi
Dr. Rajesh Bansode
Mr. Vikas Kaul
Mrs. Shiwani Gupta
Dr. Anand Khandare
Mr. Aaditya Desai

COPY RIGHT

Notice to Past Authors of Multicon-W Published Articles
TCET intends to create a complete electronic archive of all articles and or other material previously published by TCET, if you have written a work that was previously published by TCET in any journal or conference proceedings prior to 2015, and you do NOT want this work to appear in the TCET digital Library, please inform multiconw@thakureducation.org, stating the Title of the work, the author(S), and where and when published.

ISBN Numbers
ISBN: 978-1-7329562-6-1
Printed in India
Thakur College of Engineering & Technology, Mumbai India

PREFACE

It is a great honour for us to write a message for the INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING AND NETWORKING (IC-ICN 2020) scheduled during Feb.28-29, 2020.

This is the eleventh event in the series of international conferences organized by TCET under the umbrella of MULTICON since the first event ICWET 2010. This event is organized in view of strengthening the research culture among its stakeholders namely students, faculty and Industry. Moreover, need of the time is technological development in the areas of Intelligent Systems and Communication Engineering which can simplify our life in the eco-friendly environment with better connectivity and security. The 11th Annual International Conference, IC-ICN 2020 has affiliation with Scopus Indexed journal for Intelligent systems, leading publication house Springer, McGraw Hill, and Conference proceeding with ISBN number. It serves as a premier platform that gathers all academicians, researchers and professionals, in the relevant engineering disciplines and domains, to showcase their research contributions and share their thoughts. IC-ICN 2020 has gained wide publicity through website, social media coverage, and visits to various colleges by the team of faculty members and our well-wishers. As a result, this event has got overwhelming response.

TCET has strong belief in quality and relation building. Therefore, a lot of care has been taken for branding the event, identifying resources, logistic support required for event, compilation and printing of conference proceeding, souvenir etc. Effort has been taken to make the delegates feel at home though away from home with confidence of doing something for development of nation.

TCET has successfully held in the past ten international events in a row. ICWET 2010, 11, 12 and Multicon-W 2013, 14, 15,16,17,18 and 19 in association with International journals of repute and multinationals in the IT industry with an objective to reach a new horizon. We feel that the present endeavour could not have been possible without moral support, strong belief and time to time motivation from management of Thakur Educational Group.

During the two days there were about 235 presentations from national as well as International Researchers and Industrial personnel, Idea presentations with deliberation by delegates and resource persons. Our best wishes and good luck to all.

We thank to all the members of organizing and editorial committee for supporting the event and extending their cooperation to make it a grand successful event.

Dr. Sheetal Rathi
Convenor: IC-ICN 2020
Dr. Rajesh Bansode
Joint Convenor: IC-ICN 2020

CONTENTS

PART-I COMPUTER ENGINEERING

COMPUTER NETWORKING

PART - II INFORMATION TECHNOLOGY

COMPUTER NETWORKING

Review: Development of Agriculture field using Machine Learning

Ms. Rupali A. Meshram[1] and Dr. A. S. Alvi[2]

[1] PRMIT&R, Badnera, Amravati, India
[2] Professor, PRMIT&R, Badnera, Amravati, India
rupalimeshram235@gmail.com

Abstract**. In India, farmers are facing more problems compare to other coun-tries because of economical condition. Actually farmers has problems like a lack of knowledge about weather, unaware of soil contents, plant diseases, in-sufficient water level, quality of seeds and traditional way of farming. Indian farmers are not aware of new way of farming. Various Machine Learning tech-niques are developed to improve agriculture field. This review article not gives the all categorized of agriculture field it mostly focuses on yield prediction, dis-ease detection, weed detection, crop quality, species recognition, water man-agement and soil management. The presented articles demonstrate how agricul-ture will benefit from machine learning technologies, deep learning and big da-ta. By applying machine learning to the problems of agriculture field; farmers will improve fruits quality and crop production. Smart farming is the need of today's farmers.**

Keywords: Machine learning, Deep learning, Big data, deep convolutional neural networks (CNNs), Support Vector Machine (SVM).

I. INTRODUCTION

Machine learning is an imminent field of computer science which can be applied to the farming sector quite effectively. Machine Learning learns from past experiences and is able to build a model which would most likely be able to comprehend future instances. Machine Learning and Big Data are two different domains there is no direct relation between them; but before training data set there is need to prepossessed the data where the big data will comes before applying algorithm. Machine Learning usually works with huge chunks of data (Big Data). Big data machine learning would be totally irrelevant, because to learn anything from data there is a need to have a large number of 'training samples' so that all possible scenarios are exhausted and also to avoid faulty training due to few erroneous datasets. One of the main reasons why big data and machine learning are used together is because big data is more likely to be a preprocessing step to machine learning. In agriculture, big data analysis presents the availability of hardware and software, techniques and methods for big data analy-sis [1]. Big Data is defined by several characteristics beyond size, particularly, the volume, velocity, variety, and veracity of the data [2].

Deep learning is a subfield of machine learning. It technically is machine learning and functions in the same way but it has different capabilities. Deep learning is based on neural network models and work well on large sample sizes rather than small sam-ple sizes. Deep learning provides high accuracy, existing commonly used image processing technique which is use for classification or prediction. The overall benefits of deep learning are encouraging for its further use towards smarter, more sustainable farming and more secure food production [3].

Bayesian Networks can be advanced [4] in three areas of agricultural. These areas are: Text Mining, Sensor Data, and Dynamic Bayesian Networks. Cheap sensor de-vices will allow future Bayesian Network designers to either: a use exclusively data-driven construction methodologies or augment manually constructed systems with data-driven techniques. Data mining techniques like artificial neural network, Baye-sian networks and support vector machines in agricultural domain was developing early. According to authors [5] there is need to understand how these techniques can be used with complex agricultural datasets and used for crop yield prediction with an integration of both seasonally and spatially using GIS technologies.

II. LITERATURE SURVEY

Various researchers provide the new way of farming by introducing machine learning techniques in agriculture field. New researchers are needed to know where to use these techniques to improve crop production. Authors [6], have been proposed Crop Selection Method (CSM) which use to solve crop selection problem, and maximize net yield rate of crop.

According to authors [7] there are three SOM based models, namely, supervised Kohonen networks (SKN), counter-propagation artificial networks (CP-ANN) and XY-fusion (XY-F) which use to predict within field variation in wheat yield, based on on-line multi-layer soil data, and satellite

imagery crop growth characteristics. The SKN model can be used to predict wheat yield and to classify field area into different yield potential zones.

Novel active learning method proposed by authors [8] that can recognise crop and weed species by using differences in their spectral reflectance. Best results for the active learning were achieved by using SOM and MOG based one-class classifiers while mediocre results were obtained using auto-encoder network and SVM based one-class classifier.

According to authors [15] extreme learning machines (ELM), multivariate regres-sion splines (MARS), M5 Tree and support vector regression (SVR) use to improve the original DRASTIC method for groundwater contamination risk mapping, in the complex aquifer system.

The smartphone based system that uses the classification model learnt to do real-time prediction of the state of health of a farmer's garden [20]. Remote sever can determine the state of disease of a plant through the uploaded image.

Authors [28] have been employed SVM using different kernel functions including Cauchy kernel, Invmult Kernel and Laplacian Kernel. Grid search and N-fold cross-validation techniques which extract relevant features related to image of tomato leaf and to detect and identify type of disease that infects tomato plant.

For an automatic crop disease recognition method the probabilistic neural networks (PNNs) classifier used by authors [30], which combined the statistical features of leaf images and meteorological data.

Principle Component Analysis (PCA) and Support vector machine (SVM) were classifying infected/uninfected tomato fruits according to its external surface; it uses feature fusion method with color and texture features [31]. The proposed system, grading tomato fruits based on surface defects.

To identify the phenological stage classification of rice crop with multitemporal co-polar X-band SAR images authors [32] were used Support vector machines (SVM) with linear and nonlinear kernel, k-nearest neighbors (kNN), and decision trees (DT).

Support vector machine (SVM) and a random forest (RF) method used by authors [33] for assessing the feasibility of in-season winter wheat mapping and investigating potential classification. It is improvement by using SAR (Synthetic Aperture Radar) images, optical images, and the integration of both types of data in urban agricultural regions with complex planting structures in Southern China.Authors [34] used Convo-lutional Neural Network (CNN) to predict the type of previously unseen images of plant seedling.

Following table gives detail about methodology used on particular problem of agriculture field with future scope provided by researchers.

Table 1. Survey on agriculture problem with machine learning.

Methodology	Purpose	Future Scope
Multidisciplinary model based on Sen (Io sor T), s, Clo ud- Computing, Mo bile - Computing, Big- Data analysis [9].	Increase in agricultural pro- duction and for cost control of Agro-products.	Focusing on interfacing different soil nutrient sensors with beagle black bone. Use various data mining algorithms for agricultural Big-Data analysis for getting the preferred outcome.
	A cost-effective and high- performance option for field-	Texture features extracted from high temporal-spatial resolution fusion data.

Deep Neural Network (DNN) [10].	level and in-season crop-type classification for corn/soybean dominated Corn Belt landscape (case study in Champaign County, Illinois).	To extending this approach and including other types of crop types and natural vegetation types. Reducing expected uncertainties in the classification approach. To use multi-sensor fusion data, such as the STARFM algorithm.
convolution 4 neural -works(CNN) [11]	Classification method for high-resolution remote sensing images.	If the multi-temporal characteristics of crop remote sensing images can be added to the CNN training, the prediction may get good results.
novel pipeline [12]	Accurately detected object and segmentation for extracting stalk counts and stalk width. With novel pipeline they demonstrated accurate measurement of multiple plants attributes.	To integrate more accurate positioning to merge multiple views of the stalks into more accurate measurements of stalk-count and stalk-width.
Least squares support machine		

hines and Cubist method compare over the two multivariate methods: principal component regression and partial least squares regression [13]	Estimating soil total nitrogen (TN) organic carbon (OC) and moisture content (MC). LS-SVM best for MC & OC and cubist for TN estimation of soil.	This machine learning techniques can be used in field spectroscopy for off-line and online prediction of the soil parameters studied in fields with similar soil type and variability.
A new time series model based on Long Short-Term Memory (LSTM) [14]	Predicting water table depth, especially in areas where hydrogeological data are difficult to obtain.	Deeper, wider and more robust LSTM-based model, in order to provide more accurate water table depth prediction worldwide. The proposed model also can be combined with other methods, like PCA and wavelet transform. Also, it can be applied to other time series prediction tasks, such as soil water change and streamflow prediction.
Smart Irrigation Decision	Estimates the weekly irrigations needs of a plantation, on	To extend and evaluate the system in plantations different than citrus and analyse the performance under

Support System (SIDSS) [16]	the basis of both soil measurements and climatic variables..	several conditions and regions. to improve the accuracy of the system the past rainfall information may be used.
Data mining Optimization techniques, -model Genetic Algorithm [17]	Optimal water allocation relative to maximizing irrigation uniformity and minimizing yield reduction.	Address other objectives, such as economical return and improved yield.
Fuzzy-Based Classification approach (FRBCS) [18]	To estimate the ripeness of tomatoes based on color.	The proposed approach can also be applied for other climacteric crops such as mango and pepper. External features other than color such as shape, size, and texture can be involved for classification of tomato. This will help in better quality evaluation.
Decision Support System (DSS) using multiclass SVM [19]	Tomato classification between healthy and diseased leaf of Grapes plants.	The accuracy of the system can be further improved by improving the training ratio. Developing a complete system; that contains the server side components with a trained model and smart mobile device-

Deep convolutional networks [21]	Classify and detect plant diseases from leaf images.	es application. It contains features like displaying identified diseases in fruits, vegetables, and other plants, based on leaf images captured by the mobile phone camera.
Convolutional neural network models [22]	Plant disease and pest detection and diagnosis using simple leaves image of a healthy and diseased plants.	The usage of the model by training it for plant disease recognition on wider land areas, combining aerial photos of orchards and vineyards captured by drones and convolution neural networks for object detection. The development of an automated pesticide prescription system that would require a confirmation by the automated disease diagnosis system to allow the purchase of appropriate pesticides by the farmers. That would drastically limit the uncontrolled acquisition of pesticides that leads to their overuse and misuse, with the consequent catastrophic effects on the environment.
Deep convolutional net	Plant identification from leaf vein	Deep learning in agriculture, in particular weeds detection and

o wo n rk a l	patterns.	identification,
[2 3] auto A mate n d multi-class clas- sifi cati on ap- p r o a c h use s Pri nci Co pal m- ponents Analysis (P C A in) addi- tio Su n pp to ort V e c t o r Mac hines	Tomato ripeness measurement and evaluation via investigat- ing and classifying the differ- ent maturity/ripeness stages.	and seeds viability tests. To apply the proposed approach on different crops, other than tomatoes, in order to automate the whole process of harvesting and detect damages to save crops. To use nondestructive/ non- invasive detection technologies of food quali- ty/maturity such as hyperspectral imag- ing systems, colorimetric, Near Infrared
6		Convolution neural networks [27]
(SVMs) and Linear Discrimi- nant Analysis (LDA) algo-rithms [24]		Partial least square regression (PLSR), ν sup-port vector re- gression (ν-SVR), and Gaus- sian process regression (GPR) methods [29]
		Spectroscopy, and non- invasive smart sensing technologies.
Deep convolu- tional neural networks (CNNs) [25]		
		To apply other deep architectures and other training algorithms, such as the restricted Boltzmann machine which achieves a better performance on object Effectively classify 10 com- recognition. Model can extend for fault mon rice diseases through diganosis. Deeper analysis of the training
Deep convolu- tional neural network [26]		

images recognition.	method with and without labeled sam-	
ples.	The model result could be further	
extended to the distributed state estima-		
tion problems for sensor networks and		
nonlinear time-varying systems.		
Counts fruits efficiently even	To count fruits in all stages; add green	
if fruits are under	shadow, fruits to the synthetic dataset. To develop	
occluded by foliage, branches,	a mobile application based on the pro-	
or if there is some degree of	posed algorithm which can be used di-	
overlap amongst fruits (toma-	rectly by farmers for yield estimation	
to).	and cultivation practices.	
To identify and classify bana-	To test more banana and plants diseases	
na diseases.	with proposed model.	
For wheat leaf rust	disease	
detection as well as evaluating	PLSR, v-SVR, and GPR need to be	
the training sample size and	tested on various	
sensors and different		
influence of diseasesymp-	varieties of wheat in order to be used in	
toms effects on the methods	the field.	
predictions.		

III. SUMMARY AND DISCUSSION

3.1 A Subsection Sample

Table 1 provides brief literature about yield prediction, disease detection, weed detec-tion, crop quality, species recognition, water management and soil management. by referring Table 1 new researchers get the knowledge about to which algorithm or methodology used by previous researchers and it also provides future scope to partic-ular problem. It is observed that the deep convolutional neural networks (CNNs) and probabilistic neural networks (PNNs) classifier used to identify plant diseases. The deep convolutional neural networks (CNNs) applied on various plants like tomato, rice, banana to identify their diseases. Also it is used to predict types of images of plant seedling, counting fruits, plant identification and exacting plant images. Whenever we need to solve a problem involving with plant images and object detection from the image in that case we can apply deep convolutional neural networks (CNNs). The Support Vector Machine (SVM) is used with other techniques for soil estimation, healthy & diseased plant leaf, ripeness of fruits, plant diseases and stages of crops or fruits. Fuzzy Rule-Based Classification approach (FRBCS) also identify ripeness of fruits. The Support Vector Machine (SVM) applied on various plants like tomato, rice, grapes, wheat. A new time series model based on Long Short-Term Memory (LSTM) used to predict water table depth, especially in areas where hydro-geological data are difficult to obtain. Extreme learning machines (ELM), multiva-riate regression splines (MARS), M5 Tree and support vector regression (SVR) for groundwater contamination risk mapping, in the complex aquifer system. Smart Irri-gation Decision Support System (SIDSS) estimates the weekly irrigations needs of a plantation. Groundwater prediction is also has a more scope to develop new algorithm to improve crop productions.

IV. CONCLUSION AND FUTURE SCOPE

4.1 Conclusion

The aim of this article is to motivate more researchers to analyze and experiment with machine learning. To apply it for solving various agricultural problems involving yield prediction, disease detection, weed detection, crop quality, species recognition, water management and soil management. The overall benefits of survey are encourag-ing for its further use towards smarter, more sustainable farming and increase crop production. This article performed a review of machine learning, deep learning and big data in agriculture. Researcher were identified and analyzed, examining the prob-lem they addressed, tools/techniques, the proposed solution and future scope. Based on this article, the researchers can be informed about which types of agricultural applications currently use by machine learning, deep learning and big data. Survey pro-vides various methodologies used on different plant/fruits so researchers can identify other plants/fruits for their research for existing problems in agriculture field. Table 1 summarizes the methodology, purpose of methodology and future scope for those particular problems in the area of machine learning, deep learning and big data.

Future Scope. Whenever we need to solve a problem which involving with plant images and object detection from the image in that case we can apply deep convolu-tional neural networks (CNNs). So researchers can find new other

problems where they can use deep convolutional neural networks (CNNs). The Support Vector Machine (SVM) is used to identify problems related with ripeness, diseases of plants/fruits based on images. Researchers can identify different plants/fruits for exist-ing problems where they can apply SVM. A Groundwater prediction is also has a more scope to develop with new algorithm to improve crop productions. Various algorithms of Machine learning can be used for water problems. Need to improve various techniques in identification of crop and weed species with Machine learning.

References

1. Andreas Kamilaris, Andreas Kartakoullis, Francesc X. Prenafeta-Boldú, ―A review on the practice of big data analysis in agriculture‖, Computers and Electronics in Agriculture 143 (2017) 23–37.
2. Keith H. Coble, Ashok K. Mishra, Shannon Ferrell, and Terry Griffin ―Big Data in Agri-culture: A Challenge for the Future‖, Applied Economic Perspectives and Policy (2018) volume 40, number 1, pp. 79–96, doi:10.1093/aepp/ppx056.
3. A. Kamilaris, F.X. Prenafeta-Boldú, ―Deep learning in agriculture: A survey‖, Computers and Electronics in Agriculture 147 (2018) 70–90.
4. Brett Drury, Jorge Valverde-Rebaza, Maria-Fernanda Moura, Alneu de Andrade Lopes, ―A survey of the applications of Bayesian networks in agriculture‖, Engineering Applica-tions of Artificial Intelligence 65 (2017) 29–42.
5. Niketa Gandhi, Leisa J. Armstrong, ―A review of the application of data mining techniques for decision making in agriculture‖, 2nd International Conference on Contemporary Com-puting and Informatics (ic3i), 2016, pp. 1-6.
6. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, ―Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique‖, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Ener-gy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015. pp.138-145.
7. X.E. Pantazi, D. Moshou , T. Alexandridis, R.L. Whetton, A.M. Mouazen, ―Wheat yield prediction using machine learning and advanced sensing Techniques‖, Computers and Electronics in Agriculture 121 (2016) 57–65.
8. Pantazi, X.-E., et al., Active learning system for weed species recognition based on hyper-spectral sensing, Biosystems Engineering (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.01.014
9. Hemlata Channe et al, ―Multidisciplinary Model for Smart Agriculture using Internet-of-Things (IoT), Sensors, Cloud-Computing, Mobile-Computing & Big-Data Analysis‖, Int.J.Computer Technology & Applications, ISSN:2229-6093 ,Vol 6 (3), 2015, 374-382.
10. Y. Cai et al. , ―A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach‖, Remote Sensing of Environment 210 (2018) 35–47.
11. Yao Chunjing, Zhang Yueyao, Zhang Yaxuan, Haibo Liu, ―APPLICATION OF CONVOLUTIONAL NEURAL NETWORK IN CLASSIFICATION OF HIGH RESOLUTION AGRICULTURAL REMOTE SENSING IMAGES‖, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vo-lume XLII-2/W7, 2017 ISPRS Geospatial Week 2017, 18–22 September 2017, Wuhan, China.
12. H. S. Baweja (B), T. Parhar, O. Mirbod, S. Nuske, ―StalkNet: A Deep Learning Pipeline for High-Throughput Measurement of Plant Stalk Count and StalkWidth‖, Springer Inter-national Publishing AG 2018, M. Hutter and R. Siegwart (eds.), Field and Service Robot-ics, Springer Proceedings in Advanced Robotics 5, https://doi.org/10.1007/978-3-319-67361-5_18 271-284
13. Morellos, A., et al., Machine learning based prediction of soil total nitrogen, organic car-bon and moisture content by using VIS-NIR spectroscopy, Biosystems Engineering (2016), http://dx.doi.org/10.1016/j.biosystemseng.2016.04.018
14. Zhang, J., Zhu, Y., Zhang, X., Ye, M., Yang, J., Developing a Long Short-Term Memory (LSTM) based Model for Predicting Water Table Depth in Agricultural Areas, Journal of Hydrology (2018), doi: https://doi.org/10.1016/j.jhydrol.2018.04.065
15. Rahim Barzegar, Asghar Asghari Moghaddam, Ravinesh Deo, ElhamFijani, Evangelos Tziritis, ―Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms‖, Science of the Total Environment 621 (2018) 697–712.
16. H. Navarro-Hellín et al., ―A decision support system for managing irrigation in agricul-ture‖, Computers and Electronics in Agriculture 124 (2016) 121–131, http://dx.doi.org/10.1016/j.compag.2016.04.003, 0168-1699/_ 2016 Elsevier B.V.
17. L. Hassan-Esfahani et al. , ―Assessment of optimal irrigation water allocation for pressu-rized irrigation system using water balance approach, learning machines, and remotely sensed data‖, Agricultural Water Management 153 (2015) 42–50, http://dx.doi.org/10.1016/j.agwat.2015.02.005 0378-3774/© 2015 Elsevier B.V.
18. Nidhi Goela, Priti Sehgal, ―Fuzzy classification of pre-harvest tomatoes for ripeness esti-mation – An approach based on automatic rule learning using decision tree‖, Applied Soft Computing 36 (2015) 45–56.
19. Harshal Waghmare, Radha Kokare., Yogesh Dandawate, ―Detection and Classification of Diseases of Grape Plant Using Opposite Colour Local Binary Pattern Feature and Machine Learning for Automated Decision Support System‖, 3rd International Conference on Sig-nal Processing and Integrated Networks (SPIN), 2016, 513-518.
20. Godliver Owomugisha, and Ernest Mwebaze, ―Machine Learning for Plant Disease Inci-dence and Severity Measurements from Leaf Images‖, 2016 15th IEEE International Con-ference on Machine Learning and Applications, 978-1-5090-6167-9/16 $31.00 © 2016 IEEE DOI 10.1109/ICMLA.2016.126, Pp 158-163.
21. Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Ste-fanovic, ―Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification‖, Hindawi Publishing Corporation, Computational Intelligence and Neuros-cience Volume 2016, Article ID 3289801, 11 pages, http://dx.doi.org/10.1155/2016/3289801.
22. Konstantinos P. Ferentinos, ―Deep learning models for plant disease detection and diagno-sis‖, Computers and Electronics in Agriculture 145 (2018) 311–318.
23. Guillermo L. Grinblat, Lucas C. Uzal, Mónica G. Larese, Pablo M. Granitto, ―Deep learn-ing for

plant identification using vein morphological patterns‖, Computers and Electronics in Agriculture 127 (2016) 418–424.

24. El-Bendary, N., et al. Using machine learning techniques for evaluating tomato ripeness. Expert Systems with Applications, (2014), http://dx.doi.org/10.1016/j.eswa.2014.09.057.

25. Y. Lu et al. , ―Identification of rice diseases using deep convolutional neural networks‖, Neurocomputing 267 (2017) 378–384, http://dx.doi.org/10.1016/j.neucom.2017.06.023 0925-2312/© 2017 Elsevier B.V.

26. Maryam Rahnemoonfar and Clay Sheppard, ―Deep Count: Fruit Counting Based on Deep Simulated Learning‖, mdpi, Sensors 2017, 17, 905; doi:10.3390/s17040905 pp. 1-12.

27. Jihen Amara, Bassem Bouaziz, and Alsayed Algergawy, ―A Deep Learning-based Ap-proach for Banana Leaf Diseases Classification‖, B. Mitschang et al. (Hrsg.): BTW 2017 – Workshopband, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017, pp. 79-88.

28. Usama Mokhtar, Mona A. S. Alit, Aboul Ella Hassenian, Hesham Hefny, ―Tomato leaves diseases detection approach based on support vector machines‖, 2015 , IEEE, pp. 246-250.

29. Davoud Ashourloo, Hossein Aghighi, Ali Akbar Matkan, Mohammad Reza Mobasheri, and Amir Moeini Rad, ―An Investigation Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement‖, 2016 IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, pp. 1-8.

30. Shi Y, Wang X F, Zhang S W, Zhang C L. PNN based crop disease recognition with leaf image features and meteorological data. Int J Agric & Biol Eng, 2015; 8(4): 60－68.

31. Noura A. Semary, Alaa Tharwat, Esraa Elhariri, and Aboul Ella Hassanien, ―Fruit-Based Tomato Grading System Using Features Fusion and Support Vector Machine‖, 2015 D. Filev et al. (eds.), Intelligent Systems'2014, Advances in Intelligent Systems and Compu-ting 323, DOI: 10.1007/978-3-319-11310-4_35, pp. 401-410.

32. Ça˘glar Küçük, Gül¸sen Ta¸skın, and Esra Erten, ―Paddy-Rice Phenology Classification Based on Machine-Learning Methods Using Multitemporal Co-Polar X-Band SAR Im-ages‖ 2016, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, pp. 1-11.

33. Tao Zhou, Jianjun Pan, Peiyu Zhang, Shanbao Wei and Tao Han, ―Mapping Winter Wheat with Multi-Temporal SAR and Optical Images in an Urban Agricultural Region‖, Mdpi, Sensors 2017, 17, 1210; doi:10.3390/s17061210, pp. 1-16.

34. Belal A. M. Ashqar, Bassem S. Abu-Nasser, Samy S. Abu-Naser, ―Plant Seedlings Classi-fication Using Deep Learning‖, International Journal of Academic Information Systems Research (IJAISR), ISSN: 2000-002X, Vol. 3 Issue 1, January – 2019, Pages: 7-14.

35. Manuel Vázquez-Arellano, Hans W. Griepentrog, David Reiser and Dimitris S. Paraforos, ―3-D Imaging Systems for Agricultural Applications—A Review‖, Sensors 2016, 16, 618; doi:10.3390/s16050618, 1-24

36. Sahoo, S., T. A. Russo, J. Elliott, and I. Foster (2017), Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S., Water Resour. Res., 53, 3878–3895, doi:10.1002/2016WR019933.

37. Anna Chlingaryana, Salah Sukkarieha, Brett Whelanb, ―Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review‖, Computers and Electronics in Agriculture 151 (2018) 61–69.

38. Seonyoung Parka, Jungho Ima, Eunna Janga, Jinyoung Rheeb, ―Drought assessment and monitoring through blending ofmulti-sensor indices using machine learning approaches for differentclimate regions‖, Agricultural and Forest Meteorology 216 (2016) 157–169.

39. Prof. K. A. Patil, Prof. N. R. Kale, ―A Model for Smart Agriculture Using IoT‖, Interna-tional Conference on Global Trends in Signal Processing, Information Computing and Communication, 2016, pp. 543-545.

40. Jan Behmann, Anne-Katrin Mahlein, Till Rumpf, Christoph Ro¨mer, Lutz Plu¨mer, ―A re-view of advanced machine learning methods for the detection of biotic stress in precision crop protection‖, Springer Science+Business Media New York 2014, Precision Agric, DOI 10.1007/s11119-014-9372-7.

41. VAN EVERT FK, FOUNTAS S, JAKOVETIC D, CRNOJEVIC V, TRAVLOS I & KEMPENAAR C. (2017), Big Data for weed control and crop protection. Weed Research 57, 218–233.

Categorization of Plant Leaf using CNN

Suhasini Parvatikar,Dr.Deepa Parasar
Computer Engg. Department, SCOE,Kharghar
CSE, ASET,Amity University,Panvel
er.suha@gmail.com,dparasar@mum.amity.edu

***Abstract*— As plants are main part of environment. There are variety of trees in this environment ,it is unable to recognise different kinds of plant leaf . Thus, this paper studies about plant leaf detection using CNN method. This method will help us to classify leaf, which would indirectly help us to classify different kinds of plants. Combination of deep neural networks forms a Convolutional Neural Network(CNN), which is used for recognition of plant leaf .In neural networks, Convolutional neural network mainly used for recognition of images, classifying different types of images , face recognition etc., CNN image takes image as an input , process that image and classifying it under different type of category.**

Keywords—Neutral Networks, CNN, pooling, leaf classification.

I. INTRODUCTION

There are billions of plants in the environment. Imagination of life on earth without plants is not possible, they are our source of oxygen and also balances our ecosystem. It is very difficult for a scientist,botanist or expert to recognise a leaf of a plant from such a large volume of plants. For this , we require some effective classification algorithms. Machine and deep learning technique could help to solve problems by recognizing the leaf of a particular plant.As deep learning technique performs feature extraction process automatically. We could use Deep Learning method for simplifying our process. Thus, as we are taking image as an input, we propose a method called CNN (convolutional neural network) for identifying leaf of a plant.

II. LITERATURE SURVEY

Wei Liu, Zhifang Li, Jialun Lin, Danning Liang [1], designed a system for the classification of the plant. The following steps are carried out: the images are preprocessed before giving it to classifier. Using the classification algorithm SVM (Support Vector Machine). The leaves are then classified. Next, the parameters of the SVM design are optimized using the PSO algorithm (Particle Swarm Optimization), which is then used to design and optimize the leaf recognition and classification algorithm.

Francis Rey F. Padao [7], authors here used naïve Bayes classification algorithm for identifying plant leaf. Texture and shape of leaf is given as input to the classifier. The output of naïve Bayes algorithm shows high accuracy of the model.

Prof. Meeta Kumar et al. [10], is survey paper. The researcher contrasts different classification strategies such as k-Nearest Neighbor Classifier, Probabilistic Neural Network, Genetic Algorithm, Support Vector Machine, and Principal Component Analysis. The aim of this survey is to provide knowledge of different techniques of classification which can be used to classify the leaf of a plant.

Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto and Paulus Insap Santosa [11], designed a system using Principal Component Analysis (PCA) is the technique for increasing variable space dimensionality by describing uncorrelated variable space. The goal of authors is to improve grade of leaf identification system. Shape, color, texture of leaf is given as input to system. Using the identification system, the PCA algorithm was used to transform the features into orthogonal features and the results were then given to the classifier using Probabilistic Neural Network (PNN). PCA's output increases the accuracy on datasets of the leaf identification system.

Samuel E. Buttrey, Ciril Karo [13] designed a system using a hybrid(composite) classifier. The authors have used two classifiers by using k-nearest neighbor (kNN) and classification trees. The inputs are given first to classifier to extract features and then classify them using k-NN which reduced processing load associated with k-NN, and the result shown where better than individual algorithm.

III. PROPOSED METHODOLOGY

The Proposed workflow is given below Figure 1.

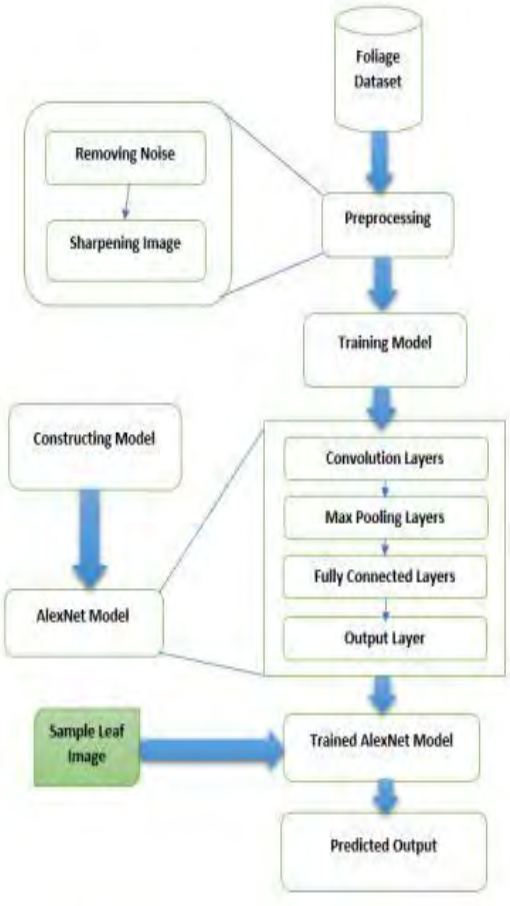


Figure 1: Flow Diagram of Proposed System

The proposed system uses CNN to classify the leaf. This application uses the specific features of each leaf which are computed and then further used for classification of them.

IV. CONVOLUTION

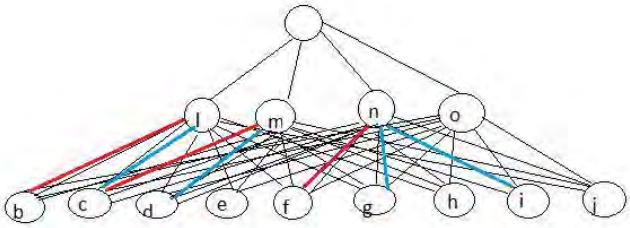


Figure 2: Fully connected network

A CNN can be implemented[19] as a feedforward neural network wherein only a few weights are active (in color) and the rest of the weights (in black) are zero. The working of CNN and information flow is as shown in Figure 2.

V. CONVOLUTIONAL NEURAL NETWORK

The use of CNNs for [18] deep learning has become increasingly popular as CNNs remove the need for manual extraction of features — the features are learned directly from CNN. Convolutional neural networks work as Neural Network in which each neuron receives some input. The only difference is the input provided to a CNN is in the form of image or a time series data on which each neuron performs operations and generates an output.

Input Layer >> Hidden Layers >> Output

Here, input layer takes the input and output gives the desired output. That means hidden layer is doing the feature extractions work as shown in Figure 3[3].

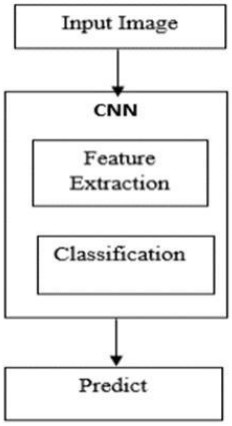


Figure 3: System Composition

The steps in CNN can be summarized [15] as:

a. Input image is given to Convolutional layer.
b. Filters / feature maps are applied to the image, which gives us a convolutional layer.
c. Apply rectifier function.
d. Perform pooling to reduce dimensionality size.
e. We then flatten our pooled feature map before inserting into an artificial neural network i.e. fully connected network.
f. Output the class using an activation function and classify images.

Step 1: Convolution: The input image is multiplied with every element by filter matrix to produce a output to reduce the size of input image. Only main features that are important are considered in this step.
Example:7 x 7 image matrix convolution is multiplied by 3 x 3 filter matrix which gives output called "Feature Map" as shown in Fig 2[17].

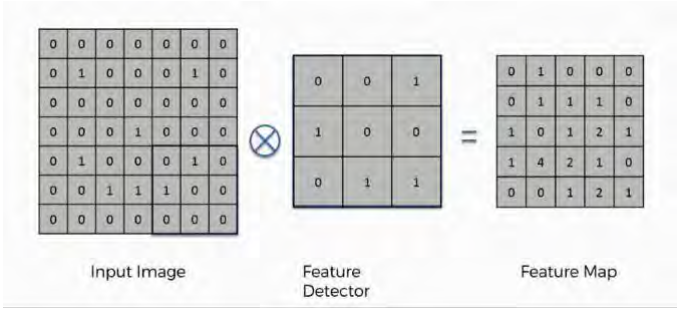


Fig 2: Convolution Process

Step 2: Apply Rectifier function**:** In NN, we use activation function to transform summed weighed Input into output. RELU is a nonlinear activation function which allows complex relationship in data to be learned **.**

This function returns 0, for negative input and x for positive value. It returns that value back. So, it can be given as:

$$f(x) = \max(0, x) \quad \text{..... eq(1)}$$

Step 3: Pooling: Its function is to gradually reduce the number of parameters and network computation. The pooling layer operates independently on each feature map. It takes a function map of each matrix of 2x2 and selects the value present in that box. Then, 2x2 matrix is passed through the entire feature map choosing the value in each pass from left to right. These values then form a pooled map of features a new matrix.

Different categories [17] of Pooling:

- Max Pooling: Maximum value is picked
- Average Pooling: average value is picked

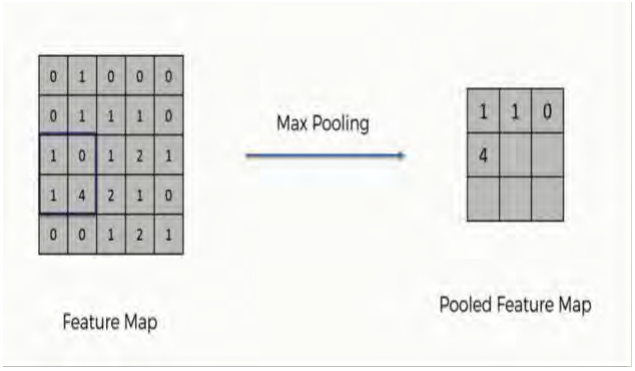


Fig 3: Pooling

Step 4: Flattening: Flattening[20] transforms the entire matrix of the pooled feature map into a single column that is then fed into the processing neural network.

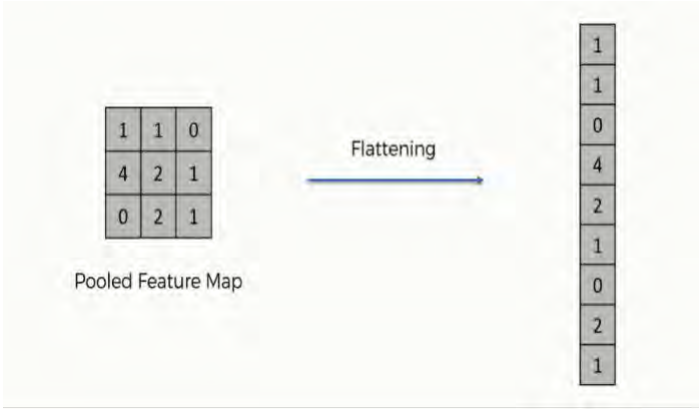


Fig 4: Flattening

Step 5: Full Connected Network: Fully connected layers[21] are an essential component of Convolutional Neural Networks (CNNs), which have been proven very successful in recognizing and classifying images for computer vision. The data will be transmitted through the network and the prediction error will be measured. It is done by the using the function as shown in eq (2):

$$f_j(z)_j = \frac{e^{z_j}}{\sum_{k=1} e^{z_k}} \quad \text{.....eq (2)}$$

The error is then propagated back to boost the prediction through the process. It is important that output produced by output layer must be scaled between 0 and 1 which represents probability of each class.

VI. DATASET

The foliage Dataset is taken from Flaviasourcsfourge.net with 32 different classes. Each of the images in the used datasets here are preprocessed in order to generate standardized input data for the training process of the network.

VII. CONCLUSION

In this paper, we studied a new method to classify leaves using the CNN model. We studied layers of simple CNN model. There are many architectures of CNN like AlexNET, GoogleNet, LeNet etc. with little difference in architecture. CNN is mostly used for input images. Thus, in fully connected network accepts weights from each node while in convolutional layer each neuron is only connected to a nearby neurons only which helps to extract only relevant information.

References

[1] Wei Liu, Zhifang Li, Jialun Lin, Danning Liang," The PSO-SVM-based Method of the Recognition of Plant Leaves", 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC),pg no 1350-1355, Haikou, China.

[2] Min Gao, Lang Lin, Richard O. Sinnott," A Mobile Application for Plant Recognition through Deep Learning", 2017 IEEE 13th International Conference on eScience,29-38, Australia.

[3] Wang-Su Jeon and Sang-Yong Rhee," Plant Leaf Recognition Using a Convolution Neural Network", International Journal of Fuzzy Logic and Intelligent Systems Vol. 17, No. 1, March 2017, pp. 26-34.

[4] Boranùekerolu, Yucelnan," Leaves recognition system using a neural network", 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30, Vienna, Austria.

[5] Timothy J. Jassmann, Rahman Tashakkori, and R. Mitchell Parry," Leaf Classification Utilizing a Convolutional Neural Network", Proceedings of the IEEE SoutheastCon 2015, April 9 - 12, 2015 - Fort Lauderdale, Florida.

[6] Sue Han Lee, Chee Seng Chan, Paul Wilkin, Paolo Remagninoz, "Deep-Plant: Plant Identification With Convolutional Neural Networks", https://arxiv.org/abs/1506.08425,submitted by 28th june 2015.

[7] Francis Rey F. Padao and Elmer A. Maravillas," Using Naïve Bayesian Method for Plant Leaf Classification Based on Shape and Texture Features", 8th IEEE International Conference Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), The Institute of Electrical and Electronics Engineers Inc. (IEEE) – Philippine Section,9-12, December 2015, Water Front Hotel, Cebu, Philippines.

[8] Chaoyun Zhang, Pan Zhou, Chenghua Li, Lijun Liu," A Convolutional Neural Network for Leaves Recognition Using Data Augmentation", 2015 IEEE International Conference on Computer and Information Technology.

[9] Sapna Sharma, Dr. Chitvan Gupta," Recognition of Plant Species based on leaf images using Multilayer Feed Forward Neural Network", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 6, Volume 2 (June 2015).

[10] Prof. Meeta Kumar, Mrunali Kamble , Shubhada Pawar , Prajakta Patil ,Neha Bonde,"Survey on Techniques for Plant Leaf Classification" International Journal of Modern Engineering Research (IJMER) 2013 Vol.1, Issue.2, pp-538-544 ISSN: 2249-6645.

[11] Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto and Paulus Insap Santosa, "Performance Improvement of Leaf Identification System Using Principal Component Analysis", International Journal of Advanced Science and Technology Vol. 44, July 2012.

[12] Jyotismita Chaki, Ranjan Parekh," Plant Leaf Recognition using Shape based Features and Neural Network classifiers", International Journal of Advanced Computer Science and Applications (IJACSA) , Vol. 2, No. 10, 2011.

[13] Samuel E. Buttrey, Ciril Karo "Using k-nearest-neighbor classification in the leaves of a tree", Computational Statistics & Data Analysis 40 (2002) pg-27 – 37 (www.elsevier.com/locate/csda)

[14] https://heartbeat.fritz.ai/a-beginners-guide-to-convolutional-neural-networks-cnn-cf26c5ee17ed

[15] https://www.superdatascience.com/convolutional-neural-networks-cnn-softmax-cross-entropy/

[16] https://github.com/Kulbear/deep-learning-nano-foundation/wiki/ReLU-and-Softmax-Activation-Functions

[17] https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

[18] mathworks.com/solutions/deep-learning/convolutional-neural-network.html

[19] mitesh Khapra nptel videos https://www.youtube.com/watch?v=PmZp5VtMwLE

[20] https://heartbeat.fritz.ai/a-beginners-guide-to-convolutional-neural-networks-cnn-cf26c5ee17ed

[21] https://missinglink.ai/guides/convolutional-neural-networks/fully-connected-layers-convolutional-neural-networks-complete-guide/

Krishak Sahayata : Prediction Of Best Crop Yield

Dr. Mrs. Gresha S Bhatia, Urjita Bedekar, Simran Bhagwandasani, Rahul Bhatia, Pranit Naik
Vivekanand Education Society's Institute of Technology,
Mumbai, India

Abstract**. Agriculture is one of the most critical occupations practiced in our country. It is an economic sector that plays an important role in the overall development of the country. Thus, the modernization of agriculture is significant and thus will lead the farmers of our country towards profit. Earlier sowing of crops was performed by considering the farmer's knowledge in a particular field and crop. However, as the weather conditions change very rapidly, farmers cultivate more and more crops that do not give an expected yield. Being this as the current situation, many of them do not have enough knowledge about the new crops and are not entirely aware of the benefits they get while farming them. Also, farm productivity can be increased by understanding and forecasting crop performance in a variety of environmental conditions. Thus, the proposed system takes the location of the user as an input. From the location, the nutrients of the soil, such as nitrogen, phosphorous, potassium is obtained. The processing part also takes into consideration two more datasets i.e. one obtained from the weather department, forecasting the weather expected in the current year, and the other data being static. This static data is the crop production and data related to the demands of various crops obtained from various government websites. The proposed system applies Machine Learning algorithms to identify the pattern among data and then process it as per input conditions which in turn, will propose the best feasible crops according to given environmental conditions. Thus, this system will only require the location of the user, and it will suggest several profitable crops providing a choice directly to the farmer about which crop to cultivate. Since historical data is also taken into account, the output is expected to be much more accurate.**

Keywords. Machine Learning, Yield Prediction, Indian Agriculture.

I. INTRODUCTION

Agriculture in India has a full-size history. More than 60 % of the land in the country is used for agriculture to cater to the needs of 1.2 billion people. Nowadays, India is ranked second worldwide in farm output. Agriculture and other sectors like forestry and fisheries accounted for 16.6 percent of GDP in 2009 [1][2]. The monetary contribution of agriculture to India's GDP is regularly declining. The production of crops relies on different factors like climatic, geographical, and organic elements. Accurate statistics about the character of an ancient yield of the crop is a critical modeling input, which is useful to farmers and authorities organization for decision-making techniques in establishing the right policies associated with subsequent manufacturing. The advances in computing and information storage have mainly provided at the maximum of information. The project has been to extract expertise from these statistics mining that may be able to bridge the understanding of the facts to the crop yield estimation. This task aimed at statistics mining strategies and follow them to the various variables consisting inside the database to set up if significant relationships may be discovered and the usage of fuzzy common sense to discover the circumstance of crops on a diverse situation of climatic conditions.

II. LITERATURE SURVEY

Literature survey is the most important part which guided us in our research. We had a look at previous papers aiming same objectives which we have and on the basis of study we plan to improve the limitations, challenges faced and various drawbacks analysed from those papers. Referring to the work done by Kunal Teeda, Nandini Vallabhaneni and Dr. T.Sridevi, [3] we understood that climatic analysis based on location, rainfall, soil type, yield capacity, irrigation techniques is a fundamental prerequisite that needs to be carefully implemented to plan better farming structures for a satisfactory yield. Farming risks have greatly reduced with the advent of technologies. In order to predict the weather parameters, empirical approaches like regression, artificial neural network, fuzzy logic and institution approach of statistics have been employed. For prediction of rainfall, clustering and classification methods of data mining have been used. The work done by D.S Zingade [4] is spectacularly presented. They have considered significant parameters mainly soil contains, water level and temperature. For the soil measure they have considered nitrogen, phosphorus and potassium along with the pH value of the soil. Taking Zinc and Sulphur nitrate into consideration along with these can help to improvise the system. Moreover, we have chosen to perform classification using random forest classifier which overfits the linear regression used above, since the parameters are quite significant in number. Analysing the Farmer's handbook published by Desai Fruits, [5] we have an understanding of choosing the sufficient parameters required for selecting the type of crop suitable for the cultivation. Project by P Priya, published in International Journal of engineering science and research technology have performed the prediction of crop yield Focusing only Tamil Nadu district[6].

We Intend to build a system that satisfies majority places of possible farming in India. Random Forest Algorithm has worked with great accuracy amongst all the other algorithms.

III. CHALLENGES IN EXISTING SYSTEMS

IoT based Plants : For constant analysis of the fields and crop growth, many projects have obtained to install IoT devices for data gathering, which are way costlier and require regular maintenance.[7]

Web Interface : Many projects have their UI based on python which are hosted on web server, our Mobile application is very much user friendly and we intend to bring the information available in the locale languages which are region specific.[8]

Parameters Under Consideration : The existing projects have a limited number of parameters considered for training. We intend to consider majority of parameters which are accountable for the plants growth. [9]

IV. PROPOSED PROBLEM STATEMENT

Accurate yield estimation would be essential for farmers to sow the crops accordingly. Machine Learning Systems are being widely used in building decision support tools for the contemporary farming system to improve yield production while reducing operational cost and environmental impact. The capability of Machine Learning algorithm to access tons of data and mold them into relatable format helps in predicting the future results when a particular test case is provided to the trained system, thereby it would provide a great support in predicting the estimated yield of a particular crop provided with sowing area, time, climatic conditions of that particular crop. Our system proposes to use an algorithm which will work accurately even when the number of parameters under consideration are more.

V. PROPOSED SYSTEM

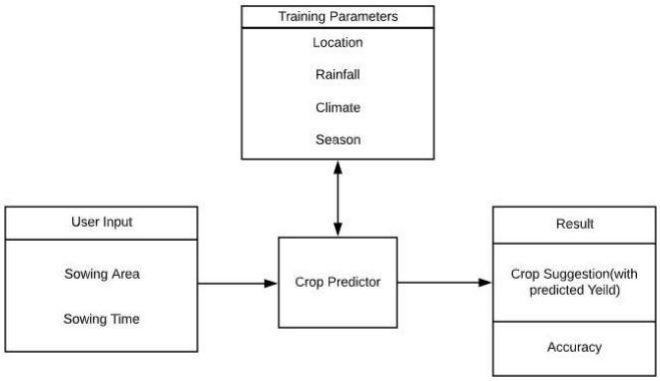
Since other systems developed are using Data Mining techniques to predict the optimal crop yield, our system aims to provide an accurate yet effective result based on more number of parameters dependent on climatic conditions to provide maximum profit, reduced risks and more than expected yield to the farmers. The proposed solution is represented in figure 1 and is described as follows :

Input : The prediction of the crop is dependent on various factors such as weather and past crop production to predict the crop correctly. The location of the user is taken as an input, which helps to get the weather conditions like humidity, temperature.

Data Acquisition : Depending on user location, the system mines the weather conditions like humidity, the temperature in the respective area from the other datasets. In a similar approach, earlier crop yield is extracted from previous data. This would give an excellent idea about the demands of crops in the location

Output : The random forest algorithm creates a forest with a number of trees. The most suitable crop is predicted by the system using the Random Forest Algorithm, and the user is provided with the best-suited crop for a given set of climatic conditions.

Fig .1. BLOCK DIAGRAM OF FARMER'S GUIDE



Sowing area and sowing time will be entered by user as parameters. Our classification algorithm will be analyzing climatic conditions such as rainfall based on the location and sowing time. Algorithm will then suggest best crop to plant for getting the maximum yield and this can be used by farmers to acquire maximum profit as observed in figure

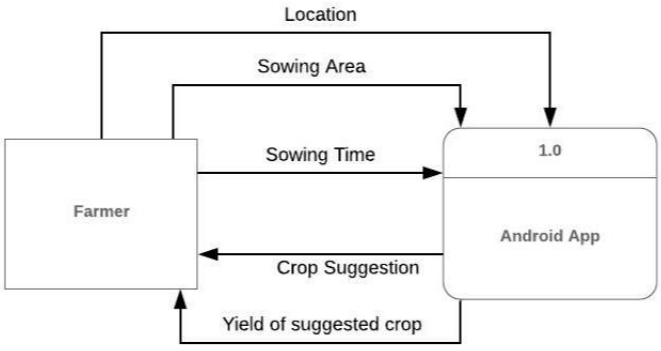


Fig .2.DATA FLOW DIAGRAM

The farmer provides location access to the android application and also provides the inputs such as the sowing area and expected sowing time. The Android Application then suggests the most suitable crop along with the expected yield.

VI. EVALUATION MEASURES

Accuracy :Accuracy can be defined as the degree to which the result of a measurement, calculation, or specification conforms
to the correct value or a standard. The Accuracy for a system is defined with the formula given as : (TP / TN) / (P + N).[10]

Error Rate :Error rate refers to the frequency of errors occurred, defined as "the ratio of total number of data units in error to
the total number of data units transmitted." The Error Rate can be calculated by using the formula : (FP+FN)/(P+N) or (1 - Accuracy).[11]

Sensitivity : Sensitivity measures the proportion of actual positives that are correctly identified as such. The sensitivity of a
system can be defined using the formula : (TP)/(P)[12].

Precision : Precision or positive predictive value, is the fraction of relevant instances among the retrieved instances. The
precision of a system can be calculated using the formula : (TP)/(TP+FP)[13].

VII. OUTPUTS OBTAINED

The parameters that we have selected to evaluate our system include accuracy, error rate and precision. Since the system is being evaluated on parameters like location, rainfall, season and climate, the quality of being precise is about 87.88% for the system developed. Having an error rate of 12.12% in a dataset of 500 trees .

VIII. CONCLUSION AND FUTURE WORK

The proposed system takes the data related to weather, and past year production and suggests which are the most suitable crops that can be cultivated in environmental conditions of various parts of India. As the system lists out the best-possible crop, this would help the farmer in the decision making of which crop is appropriate to cultivate. Since this system takes into consideration past production, it will help the farmer get insight into the demand of crops cultivated at the stipulated time in the market. The main aim of this research paper is to provide a methodology so that it can perform descriptive analytics on crop yield production in an effective manner. The future scope of the project is to collect soil nutrients for every particular piece of land and combine the trained system and data about soil nutrients to get a cumulative output on the basis of all these factors.

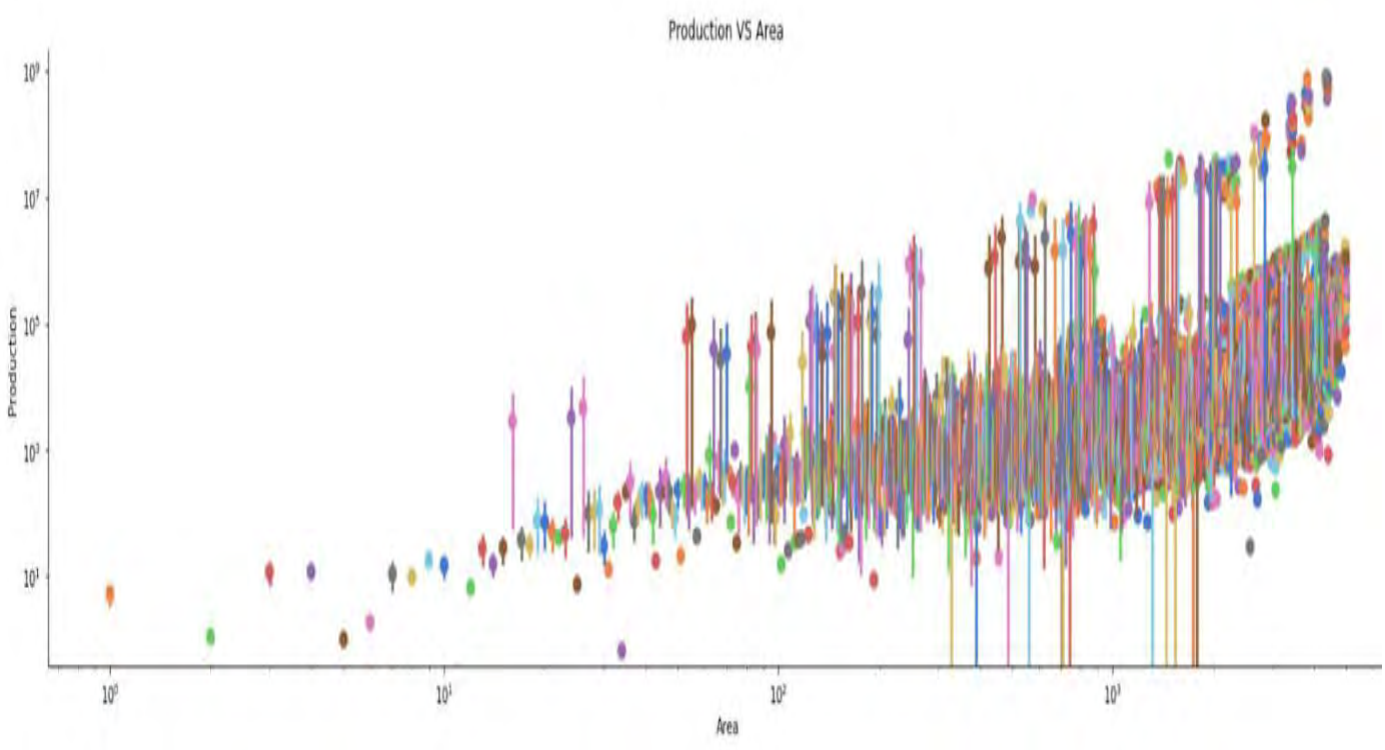


Fig. 3.Plot for Yield vs Area(Insight from data

References

[1]http://www.expo2015.org/magazine/en/economy/agriculture-remains-central-to-the-world-economy.html [2] https://en.wikipedia.org/wiki/Agriculture_in_India

[3]Analysis of Weather Attributes to Predict Crops for the Season Using Data Mining,Kunal Teeda1 Nandini Vallabhaneni2 Dr.T.Sridevi3

[4]Machine Learning based Crop Prediction System Using Multi Linear Regression, Prof D.S Zingade

[5]https://www.manage.gov.in/publications/farmerbook.pdf

[6]PREDICTING YIELD OF THE CROP USING MACHINE LEARNING ALGORITHM,P.Priya, U.Muthaiah & M.Balamurugan

[7]https://www.iofficecorp.com/blog/cost-of-iot-sensors [8]https://www.quora.com/How-many-Indian-farmers-use-the-internet

[9]https://www.manage.gov.in/publications/farmerbook.pdf

[10]https://www.researchgate.net/post/How_can_I_calculate_the_accuracy

[11]https://classeval.wordpress.com/introduction/basic-evaluation-measures/

[12]http://gim.unmc.edu/dxtests/reviewof.htm [13]http://gim.unmc.edu/dxtests/reviewof.htm

A Fuzzy Expert System for Malaria Disease Detection

Ms. Swati Jayade1 Dr. D. T. Ingole,Prof. Mrs. Manik D. Ingole
Department of Electronics & Telecommunication, Govt. Poly., Washim, India
Department of Electronics & Telecommunication, Prof. Ram Meghe Institute of Technology
Research, Badnera, Amravati India.
swatijayade22@gmail.com, dtingole@gmail.com,manik.ingole2031964@gmail.com

Abstract.
It is found that the malaria disease is a prime and major cause to the human health. The pernicious trappings of malaria stooge to the physical body can't be making light of it. In this research work, a fuzzy based expert system for the total handling of malaria disease had granted for providing judgment support platform to the specialist and healthcare researchers in the same endem-ic province. The proposed and implemented system consists of major compo-nents which include the Cognitive content, the Fuzzification, the Inference en-gine and De-fuzzification for decision making. The fuzzy inference engine de-veloped during this work is that the Root Sum Square. This method is the de-piction of inference that was designed and developed to infer the info from the fuzzy based rules used in this algorithm. Triangular fuzzy membership function was accustomed that shows the degree of attendance of every input specifica-tion and therefore the de-fuzzification technique employed during this research is that the centre of gravity. This fuzzy based expert system had been developed to help and to support clinical perception, diagnosis and therefore the expert's proficiency. For validation and empharical analysis the data of thirty patients with malaria defection was used. The results that were calculated are within the range of that was predefined and predicted by the territory proficient.

Keywords: Malaria, Fuzzy Logic, Cognitive Content, Fuzzy Expert System.

1 Introduction

The ruinous accouterments's of malaria parasites to the human being as well as other animal cannot be undervalue. Malaria is a harmful disease that caused principally by plasmodium falciparum. P. falciparum and three alternative malaria bedbugs that taint human (p. vivax, p.ovale, and p. malariae) are impart by many type of species of Anopheles mosquitoes widely in the tropics [1][2][3]. Malaria disease is that the

fa-miliar explanation for mortality within the tropics. Today's nature is one beside ex-panding approach to intelligent systems. In new time, Artificial Intelligent methods have undoubtedly been utilized in medical applications and research exercise are po-tent on doctor systems as reciprocal solution to standard technique for locating solu-tion to medical problems [10] [12]. The evolution of data technology has opened un-paralleled convenience in health healthcare delivery system because the demand for intelligent and knowledge-based systems has heightened as contemporary medical practices now a day become highly knowledge-intensive. The diagnosing of regional diseases embroils several steps of hysteria and imprecision [11].The assignment of health problem diagnosis and handling it is compound due to the various functionality involved. It's difficulty level is high due to tons of imprecision and uncertainties. It is not possible for patients to illustrate indeed how they feel and sense, medical practi-tionars and nurses cannot understand exactly what they identify or observe, and la-boratories finding are dotted with some possible errors caused may be by the care-lessness of lab workers or improper functioning of the equipments in lab.

Medical workers many not squarely identify how diseases transform the traditional working of the body [12]. Out of these ramifications in practice make long-established quantita-tive path of study disproportionate. Computer workstation tools help to arrange, stock up and reclaim relevant medicinal information desired by the practitioner in handle with each severe case and suggesting appropriate diagnosis, prognosis, remedial deci-sions and decision-making procedure [13]. A specialist framework is a savvy PC code that uses the information storehouse of at least one specialists and conclusion methods for issue taking care of

[14]. Human specialists take care of issues by utilizing a union of genuine information and thinking capacity. In a specialist framework, proposals two imperative are contained in two separate however related segments: an informa-tion base and an induction motor. The information base get ready explicit realities and rules about the subject and the derivation motor set up the thinking capacity that empowers the master framework to shape result. Master frameworks additionally give advantageous apparatuses as UI and elucidation offices. UIs, similarly as with any application, empower individuals to shape doubt, furnish data and connect with the framework. The application and use of fuzzy based perception to medical review and diagnosis of diseases are reviewed in these literatures [13] [16][17]. This experimen-tation work represents a fuzzy expert system for the administration of malaria disease. It was developed by considering clinical perception, medical diagnosis and the do-main experts. The main aim of this work is to provide a decision backing stack to scientists working on and other medicinal healthcare practitioners. This system will help medical practitioners for diagnosing and decision making in case of malaria where there is a shortfall of doctors in the benefit of society.

2 Literature Review

The symptomatic choices taken by restorative specialists rely on recognition, expe-rience, aptitude, information, capacity, and impression of the therapeutic researcher. As the unpredictability of framework builds, it is difficult to pursue a specific way of finding with no error. Fuzzy-based rationale introduces incredible thinking strategies that can deal with vulnerabilities and imprecision. Numerous frameworks have trans-formed into noteworthy in the extension and continuance of the human services zone. As of late, much research exertion has been fixed in structuring knowledge frame-works. A few reciprocal works have indicated that tropical sicknesses garbage a sig-nificant general wellbeing contest in the tropics [4] [5]. Be that as it may, the delibe

rate endeavor is constantly been made to control the acceleration and transmission of tropical ailments inside and between neighborhoods. In the work vehicle tried out by [6], it was accounted for that month to month jungle fever rate rates and vector densi-ties were utilized for reconnaissance and adaptable tuning of the ecological adminis-tration procedures; which brought about an elevated level of work. Inside 3-5 years, intestinal sickness related casualty, grimness and commonness rates were diminished by 70-95%. In an ongoing report, it was inferred that the jungle fever control program that underline natural administration were profoundly viable in minishing grimness and mortality [12]
the framework was to sponsorship restorative master in the difficult and burdensome errand of diagnosing and giving treatment to tropical illnesses. The framework gave a structure that will help the therapeutic workforce in provincial regions, where there are a shortage of specialists, during the time spent commitment to essential human services to the individuals. In a similar perspective, [3] built up a fuzzy-based master framework approach utilizing various specialists for gainful follow-up of endemic infections.

In the design of LEPDIAG the significant highlights that were been itemized are a various master condition, a homeostatic capable containing the model of safe return, a presentation evaluator that can relate the watched signs and indication with those foresee by the homeostatic master and a prognostic master which improves the administration agenda for the patients. In [17] they investigate the issues of under-treatment and uncertainty condition of acute asthma cases in ED. A novel approach, known as the fuzzy logic principle is employed to determine the severity of acute asthma. The fuzzy set theory, known as Fuzzy Rule-based Expert System for Asthma Severity (FRESAS) determination is embedded into the expert system (ES) to assess the severity of asthma among patients in ED. A medical expert system for managing tropical diseases was suggested by [2]. The proposed Medical Expert Solution (MES) system was to assist medical doctors to diagnose syndrome related to a given tropical disease, suggests the likely ailment, and advances possible treatment based on the MES diagnosis.

The objective of the research was to apply the concept of fuzzy logic technology to determine the degree of harshness on tropical diseases. The root sum square of draw-ing belief was employed to ascertain the data from the rules developed.In another examination did by [13], the financial results of jungle fever control ar-ranging were featured. The utilized of master frameworks in therapeutic applications is powerful this can be found in [14] [15]. A decent number of master frameworks have been a development on tropical ailments. In [13] a specialist framework is created on tropical ailments to help paramedical staff during exercise and in the understanding of numerous basic illnesses exhibited at their facilities. In another exami-nation completed by

[14], a specialist framework on endemic tropical ailments was developed. The objective of the framework was to be a wellspring of training for spe-cialists, and for therapeutic understudies, on the analysis of a portion of the essential tropical endemic illnesses. In [8] they planned a paper to sponsorship clinical judg-ment making. The paper centers around the separation between three sorts of clinical choice help devices: for data the board, centering thought, and patient-explicit as-sessment. There are numerous master frameworks for restorative determination and remedies of some tropical tainting dependent on choices are presented in [7] [9] [10]

[11] [14]. In [15] authors have explained the seriousness of setting preference for health research. In their paper, the point was to organize inquire about in dismissed tropical illness than prominent tropical maladies. An expository strategy that depends on the prioritization system of the worldwide discussion of wellbeing research was utilized for information examination.

In [4] scientists stress the mindfulness that examination is basic in the battle against tropical illness. In any case, the tight assets empty can support just a small amount of the promising exploration space. Thus, prioritization is fundamental for wellbeing research and observable exercise has gone into creating a powerful prioritization framework [4] [12] [13]. Tropical sickness look into Special Program for Research and guidance in Tropical Diseases, was made to address the requirement for an ex-amination into underestimated tropical malady that speaks to significant general med-ical issues in advancing nations [14]. In [17] researchers broke down and assessed Decision Support Systems and Expert Systems which are utilized as help for clinical choices just as potential outcomes to redesign the osmosis and empower the more prominent compromise of these frameworks in the present arrangement of the human services. These creators went to the accompanying results: instruments that are uti-lized in the clinical choice frameworks are classified as per specialized accessibility, and in that way, their utilization is tight and particularly their absorption with new frameworks for quality medicinal services. Along these lines, the advancement of new frameworks for clinical choice help is huge. Creators in [9] built up a structure for the use of information innovation to the administration of tropical infections. The point of

3 Fuzzy Logic

As stated by authors in [16], Fuzzy-based rationale is settled as a lot of numerical standards for information portrayal dependent on degrees of investment as opposed to on fresh enrollment of traditional paired rationale. This amazing asset to accessory imprecision and vulnerability was at first acquainted by [16] with improved agree-ment, durability and ease answers for genuine issues. Fuzzy-based sets have been applied in numerous fields in which inner conflict assumes a key job. Restorative determination is a praiseworthy case of equivocalness and vulnerability. Fuzzy-based set hypothesis is an input to the interest for thoughts and approaches for dealing with non-measurable vulnerability. A fuzzy-based set is a set with fuzzy edge. Characte-rized fuzzy-based sets or classes for every factor permits transitional evaluations of participation in them, which implies each set could have components that live mostly to it; the level of having a place is called enrollment functions extending from 0 to1. If X is the Universe of talk and its components are de-noted as X , interestingly with fresh set, at that point the fuzzy-based set An of X has attributes work related to it. The fuzzy-based set is communicated by a participation work, de-fined as seek af

□□□□X□□□□□□□□□□□ □ if x is not in A

0 □□□ □ □X□□□□ 1 if x is partly added in A.

□□□□X□□ represent to which the integer value X included in the fuzzy set A. The data-value 0 coincide to the infinite non-set membership and the value 1 belongs to the total infinite membership. Therefore, a fuzzy set membership function □□□□X□□ shows the association-degree to some value element X of the domain of discourse X under consideration. This maps each value of X to a data-membership scale between 0 and 1 in various functions like Triangular, Trapezoidal, curvaceous and Gaussian. Triangu-lar fuzzy-membership function that is extensively used is selected in this study of research. Triangular fuzzy-membership value function is computes by the equations as below:

$$\mu_A(X)=\begin{cases}0 & if\ x\le a\\ \dfrac{x-a}{c-a} & if\ x\in[a,c]\\ \dfrac{b-x}{c-b} & if\ x\in[b,c]\\ 0 & if\ x\ge c\end{cases}$$

where a, b, and c have been defined by experts doctors.

3.1 Fuzzy Expert System

Many fuzzy expert systems have been developed but the efficiency is depending basically upon the decision of the domain scientists on different problems related to the work under consideration. This system is evaluated with the help of medical practitioners and specialist. The developed system, christened has an structural design conferred in following figure 1. The development of system involves many components like fuzzification, deduction (inference) engine and de-fuzzification. It is a rule based system that uses fuzzy logic rather that Boolean logic. The system is designed and developed solely based on the knowledge-base, fuzzy logic rules and dataset.

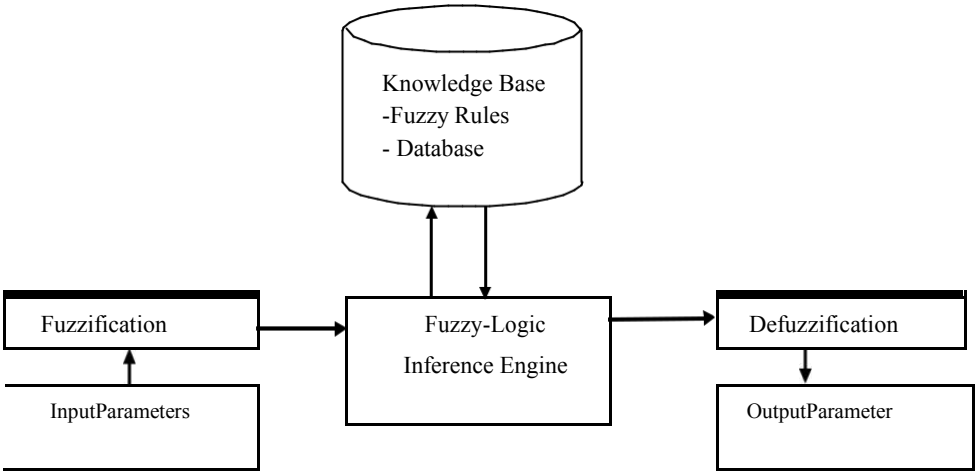


Fig. 1: Architecture of Proposed System

Procedure for Fuzzy logic based Malaria Diagnosis

The algorithm designed for the fuzzy rules based diagnostic process of malaria :

1: First give m signs and observed symptoms of patient as an input the classification.

where m = numeral of signs and syndrome.

2: Look for the knowledge-base for the disease d, which has the syndrome identified.

3: Then find the weighing factors (wf) (it is the degree of intensity of symptoms associated)

wf = 1, 2, 3, 4

Where 1 = Mild, 2 = Moderate, 3 = Severe, 4 = Very Severe.

4: Use fuzzy logic rules on the input.

5: Correlate the fuzzy data as inputs into their particular weighing factors to find out their degree of relationship.

6: Calculate the fuzzy rule base evaluating.

7: Find out the firing strength of the rules R.

8: Compute the quantity of truth R, of each rules by evaluating the nonzero minimum data value.

9: Identify the level/depth of the disease.

10: Declare the fuzzy system investigation for decision making

The Knowledge-Base of the System

Information (knowledge repository) is a key factor in the presentation of canny frameworks. The information base of the proposed framework is made out of an or-ganized and compact portrayal of the information on space specialists of tropical pre-scriptions. The unpredictable information is worried about sureness, rule, and pres-ence of tropical sicknesses, which were as often as possible settled upon by specialists in the field of tropical prescription. For the assurance of this exploration, intestinal sickness as a realized central malady is dealt with.

Fuzzification

Fuzzification process is the dispose of identifying and changing a absolute scalar value into a fuzzy set. This is accomplished with various sorts of fuzzifiers. There are commonly four kinds of fuzzifiers, which are utilized for the fuzzification procedure. They are: Trapezoidal fuzzifier, Triangular fuzzifiers, Singleton fuzzifier, and Gaus-sian fuzzifier [17]. Traingular fuzzifier which is broadly utilized will be utilized in this examination.

Fuzzification of information is completed by choosing input parameters into the flat pivot and anticipating vertically to the overhead limit of participation capacity to de-cide the level of enrollment.

The initial phase in the improvement of fluffy rationale based master framework is to develop fluffy sets for the parameters. This is appeared in conditions underneath. Based on do-principle specialists' information, both information and yield parameters chose for this exploration were portrayed with four semantic factors (mild i.e. gentle, moderate, severe :serious and very severe: extreme). The detail range of fuzzy value for all variables is given in following table 1:

FuzzyLinguistic Variables	Fuzzy RangeValues
Mild- gentle	□□□□□□□x □ 0.3
Moderate-	□□□□□□□□x □ 0.6
Severe-serious	0.6 □□□x □ 0.8
Very Severe-extreme	0.8 □ x □ 1.0

Table 1: Range of Fuzzy Set Values

Fuzzification starts with the upset of the crude information utilizing the capacities that are communicated in conditions beneath. During the activity, phonetic factors are look at utilizing triangular enrollment work and are go-to by consolidating level of participation running from 0 to 1 as appeared in conditions underneath. These recipes are unfaltering by the help of one and the other master specialists in the field of tropi-cal prescription and writing. The following stage in the fuzzification procedure is the advancement of fluffy (fuzzy) standards. The fuzzylogic principles for this examination were created with the help of area specialists. The knowledge-base of this system has so many fuzzy rules designed with the aid of consolidation theory but only the valid rules (20 in number) selected those are suggested by the field experts. Following are the some of the sample fuzzy rules for malaria.

Some of the rules (Rules 1, Rules 2, 18 and Rule 20) can be interpreted as fol-lows:

Rule1: IF fever = mild and headache = mild and nausea = mild and vomiting = mild and jaundice = mild and enlarge liver = mild and joint pain = mild and body weakness= mild, and dizziness = severe, and loss of appetite = mild and MP = mild THEN malaria = mild.

Rule 2: IF fever = moderate and headache = mild and nausea = mild and vomiting = mild and jaundice = mild and enlarge liver = mild and joint pain = moderate and body weakness = severe and dizziness = very severe, and loss of appetite = severe and MP = moderate THEN malaria = moderate.

Rule 18: IF fever = moderate and headache = Very Severe and nausea = Very Severe and vomiting = mild and jaundice = Severe and enlarge liver = Severe and joint pain = moderate and body weakness = severe and dizziness = very severe, and loss of appetite = very severe and MP = severe THEN malaria = very severe.

Rule 20: IF fever = very severe and headache = severe and nausea = severe and vomiting = moderate and jaundice = severe and enlarge liver = severe and joint pain = severe and body weakness = severe and dizziness = severe and loss of appetite = se-vere and MP = moderate THEN malaria = very severe.

Equations for fuzzification process:

$$\mu_{Mild}(X)=\begin{cases}0 & if \quad x\le 0.1\\ \frac{x-0.1}{0.2} & if \quad 0.1\le x\le 0.3\\ \frac{0.2-x}{0.1} & if \quad 0.2\le x\le 0.3\\ 0 & if \quad x\ge .0.2\end{cases}$$

$$\mu_{Moderate}(X)=\begin{cases}0 & if \quad x\le 0.3\\ \frac{x-0.3}{0.3} & if \quad 0.3\le x\le 0.6\\ \frac{0.45-x}{0.15} & if \quad 0.45\le x\le 0.6\\ 0 & if \quad x\ge .0.45\end{cases}$$

$$\mu_{Severe}(X)=\begin{cases}0 & if \quad x\le 0.5\\ \frac{x-0.6}{0.2} & if \quad 0.6\le x\le 0.8\\ \frac{0.7-x}{0.1} & if \quad 0.7\le x\le 0.8\\ 0 & if \quad x\ge .0.7\end{cases}$$

$$\mu_{Very\ Severe}(X)=\begin{cases}0 & if \quad x\le 0.8\\ \frac{x-0.1}{0.2} & if \quad 0.8\le x\le 1.0\\ \frac{0.2-x}{0.1} & if \quad 0.9\le x\le 1.0\\ 0 & if \quad x\le 1.0\end{cases}$$

4 Experimental Results

In the experimental analysis total 30 patient's data is validated. It is observed that all the 20 rules are fired according to the input values of symptoms narrated by the pa-tient. According to the result analysis the output of this expert system is on the same line with the domain medical expert with +_ 5% differences as shown in the graph figure 2. For further understanding one sample analysis is given here: seventeenth (17) rules were fired out for patient number 30. i.e. 17 rules generated non-zero min-imum values from the fuzzy rule base for malaria. For all of the linguistic variables: mild, moderate, severe and very severe, the respective output membership function strength (range: 0-1) from the possible rules (R1 to R20) are computed using RSS inference technique as shown in equations below:

Mild $\sqrt{R6^2 \; R10^2}$

$\sqrt{0.25^{\,2} \; 0.25^2}$

0.3536

Moderate $\sqrt{R2^2 \; R5^2 \; R8^2 \; R9^2 \; R13^2 \; R19^2}$

$\sqrt{0.25^2 \square 0.25^2 \square 0.25^2 \square 0.25^2 \square 0.25^2 \square 0.25^2}$

0 .6124

Severe $\sqrt{R3^2 \square R7^2 \square R12^2 \square R14^2 \square R15^2}$

$\sqrt{0.5^2 \square 0.25^2 \square 0.25^2 \square 0.5^2 \square 0.25^2}$

0 .8291

Very Severe $\sqrt{R11^2 \square R16^2 \square R17^2 \square R20^2}$

$\sqrt{0.5^2 \square 0.25^2 \square 0.25^2 \square 0.25^2}$ 0 .6614

The output from the fuzzylogic set calculated using the RSS later on defuzzified to get the crisp result for decision making. Then for defuzzification it uses the discrete center of gravity technique, the sample calculation is as shown in equation below:

$$= \frac{0.5 * 0.2 + 0.75 * 0.4 + (0.79 * 0.62) + (1.22 * 0.98)}{(0.5 + 0.78 + 0.795 + 1.3)}$$

$= 0.62, \quad = 62\%$

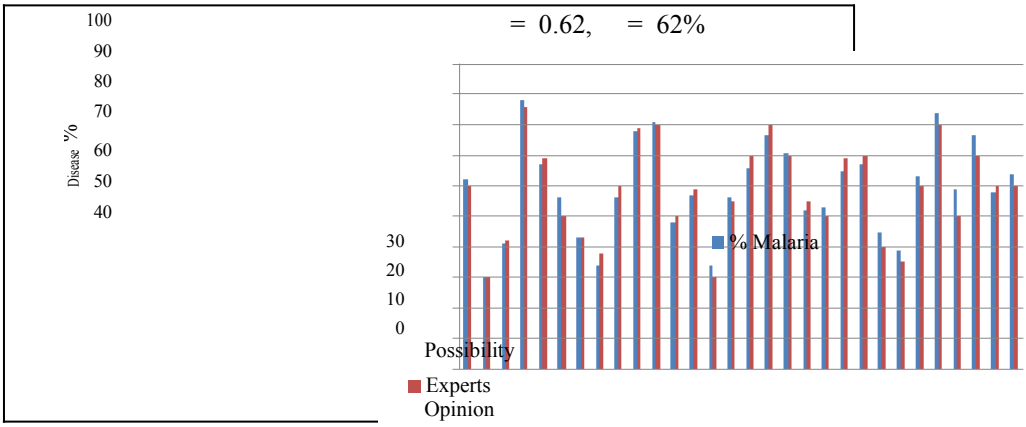


Fig2. Comparison of Fuzzy Expert systems output with Medical experts opinion

5 Conclusion

In the research work presented, the fuzzy expert system is developed for malaria dis-ease predication. The result analysis indicates that the fuzzy logic based systems can be used for direct medical diagnosis. The function of fuzzy logic for healthcare diag-nosis contributes an very helpful methodology to aid unsophisticated physicians to take a decision of malaria-disease higherside instantaneously as well as precisely. This developed system is evaluated with the domain expert's opinion and it is found that the system resembles 95.9% with the expert's decision. Hence the proposed tech-nique developed in this study, if used reasonably, could be an efficient approach for diagnosing malaria. Furthermore, this work can be scaled up as multi-disease diagno-sis system for detection other diseases showing similar kind of symptoms in patients

References

1. Matui, P., Wyatt, J. C., Pinnock, H., Sheikh, A., & McLean, S. Computer decision support systems for asthma: a systematic review. Nature Partner Journal, Primary Care Respiratory Medicine, 24, 14005. doi: 10.1038/npjpcrm.2014.5 (2014).
2. .

3. B. Apurba, K.M. Arun and B. Anupam, A fuzzy expert approach using multiple experts for dynamic follow-up of endemic diseases, Artificial Intelligence and Expert Systems, 19, 53-73 (2007).
4. C.M. Morel, Reaching maturity – 25 years of the tropical disease research, Parasitol To-day, 16, 522-528. (2000).
5. D.C. Classen, Clinical decision support systems to improve clinical practice and quality care. Journal American Medicinal Associations, 280, 180-187 (1998).
6. D.J. Briggs, A Framework for integrated environmental health impact assessment of sys-temic risks, Environmental Health. Malaria Journal, 7, 1186-1476 (2008).
7. D.L Hudson, and M.E. Cohen, Fuzzy logic in medical expert system, IEEE Eng. Med. and Bio,12, 693-698 (1994).
8. E.H. Shortliffe, Computer programs to support clinical decision making, Journal of the American Medical Association, 258, 61 – 66, (1997).
9. F.M.E. Uzoka and F.O. Famuyiwa, A Framework for the application of knowledge tech-nology to the Management of Diseases, International Journal of Health Care Quality As-surance, 17, 194- 204 (2004).
10. F.M.E.Uzoka, J. Osuji, and O. Obot, Clinical decision support system (DSS) in the diagno-sis of malaria: A case comparison of two soft computing methodologies. Expert Systems with Applications, 38,1537-1553 (2010).
11. F.M.E.Uzoka and K. Barker, Expert systems and uncertainty in medical diagnosis: A pro-posal for fuzzy-AHP hybridisation, International Journal of Medical Engineering and In-formatics, 2,329-342 (2010).
12. G.I. Doukidis, T. Cornford, and D. Foster, Medical expert system for developing countries: Evaluation in practice, Expert Systems with Applications, 7, 221-233 (1994).
13. [14] H. Devlin, and J.K Devlin, Decision support system in patient diagnosis and treat-ment. Future Rheumatology, 2,261-263 (2007).
14. Mohd Zahari, M. K., Zaaba, Z. F. Intelligent responsive indoor system (IRIS): A Potential shoplifter security alert system. Journal of Information and Communication Technology, 16(2), pp. 262–282 (2017).
15. Tan, C. F., Wahidin, L. S., Khalil, S. N., Tamaldin, N., Hu, J., & Rauterberg, G. W. M. The application of expert system : a review of research. ARPN Journal of Engineering and Applied Sciences, 11(4), pp. 2448–2453. ISSN 1819-6608 (2016).
16. Yang, C. L., Simons, E., Foty, R. G., Subbarao, P., To, T., & Dell, S. D. Misdiagnosis of asthma in schoolchildren. Pediatric Pulmonology, 52(3), pp. 293-302 (2016).
17. Mohd Sharif, N. A., Ahmad, N., Ahmad, N., Mat Desa, W. L. H., Mohamed Helmy, K., Ang, W. C., & Zainol Abidin, I. Z. A fuzzy rule-based expert system for asthma severity identification in emergency department. Journal of Information and Communication Tech-nology, 18(4), 415-438 (2019).

Data Analytics on Columnar Databases in Big Data Environment

Priyanka Bandagale,Dr. Vinayak Ashok Bharadi
IT Department
Finolex Academy of Management and Technology,,
Ratnagiri, India ,
priyanka.bandagale@famt.ac.in,vinayak.bharadi@famt.ac.in

Abstract—**Traditional approaches to data warehousing have significant drawbacks in terms of effectively delivering a solution to various businesses requirements. They were also designed primarily for Online Transaction Processing (OLTP), making them unsuited for the growing need for analytics and ad hoc queries. Significant other drawbacks include high licensing and storage cost, slow query performance against large data volumes, and difficulty in providing access to all of the data stored in multiple databases. Columnar databases can be very helpful in your big data project We have described and implemented the Column Oriented approaches for OLTP and shown how it is efficient when compared to row oriented databases.**

Keywords— OLTP, Columnar Orientation, Infobright, ICE, DPN, PHP

I. INTRODUCTION

Over the past decade, business intelligence has emerged as one of the highest priority items on Chief Information Officer (CIO) agendas. Businesses and government agencies know that mining information from the increasingly large volumes of data they collect is critical to their business or mission. Traditional database technologies were not designed to handle massive amounts of data and hence their performance degrades significantly as volumes increase. Some of the reasons for this are:

- Disk I/O is the primary limiting factor. While the cost of disk has decreased, data transfer rates have not changed. Therefore accessing large structures such as tables or indexes is slow.
- A row oriented design forces the database to retrieve all column data regardless of whether or not is it required to resolve the query.
- Load speed degrades since indexes need to be recreated as data is added; this cause huge sorts (another very slow operation). The ideal architecture to solve this problem must minimize disk I/O, access only the data required for a query, and be able to operate on the data at a much higher level to eliminate as much data as possible, as soon as possible.

A. Columnar Versus Row Databases

Column databases stores data in columns instead of rows. So when we query data the row store has to go through the whole row, whereas in column store we can just work with the specific column(s) we are interested in. This enables massive performance benefits in scaling of the database [1], [2]. Figure 1 displays organization of data on disk in column and row stores.

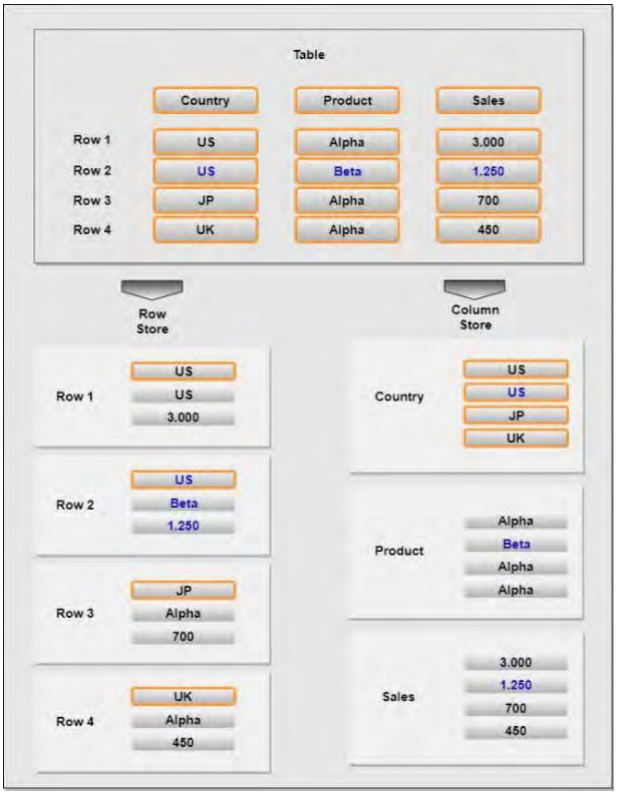


Figure 1. Actual storage on hard disk

B. Advantages Of Columnar Databases[3]

Better analytic performance: Column oriented approach allows better performance in running a large number of simultaneous queries.

b. Rapid joins and aggregation: data access streaming along column oriented data allows for incrementally computing the results of aggregate functions, which is critical for data warehouse applications.

c. Suitability for compression: Eliminates storage of multiple indexes, views and aggregations and facilitates fast improvement in compression.

d. Rapid data loading: In columnar arrangement the system effectively allows to segregate storage by columns. This means each column is build in one pass and stored separately, allowing the database system to load columns in parallel using multiple threads. Finally since the

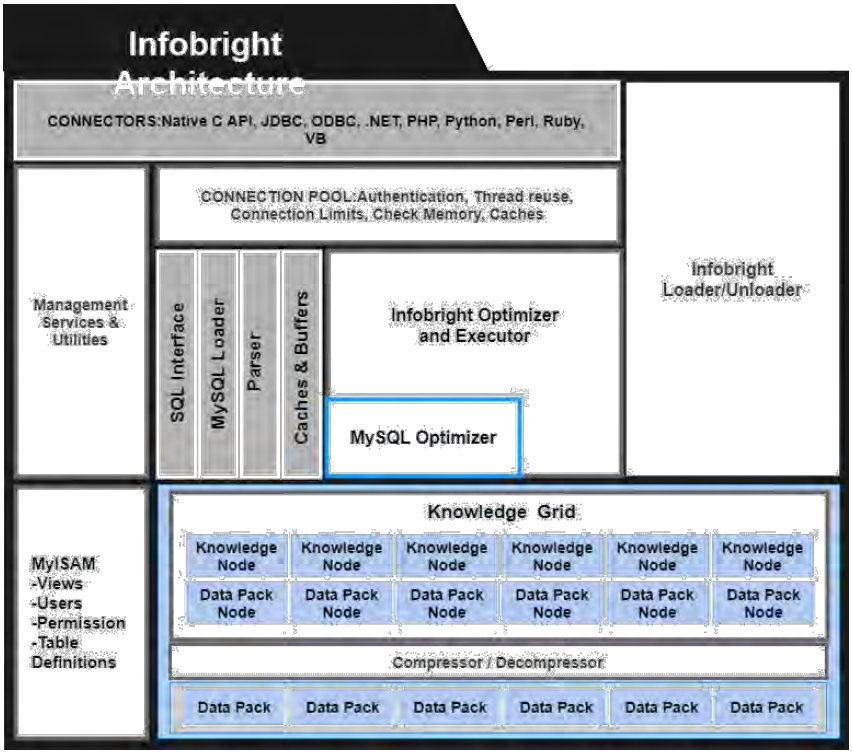


Figure 2. Architecture of Infobright

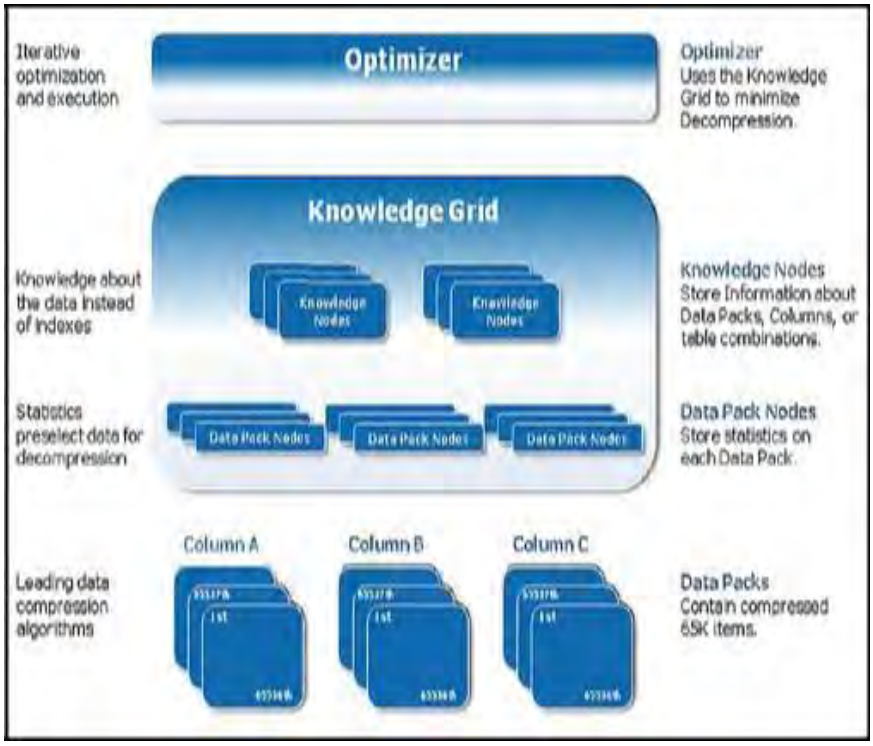


Figure 3. Data Organization

columns are stored separately entire table columns can be added or dropped without downing the system and without the need to re-tuning the system following the change.

The paper is organized in four sections followed by references. Section 2 describes the open source software which we have used to implement the system. Section 3 describes the implementation showing comparison of row and column databases. Section 4 summarizes the paper.

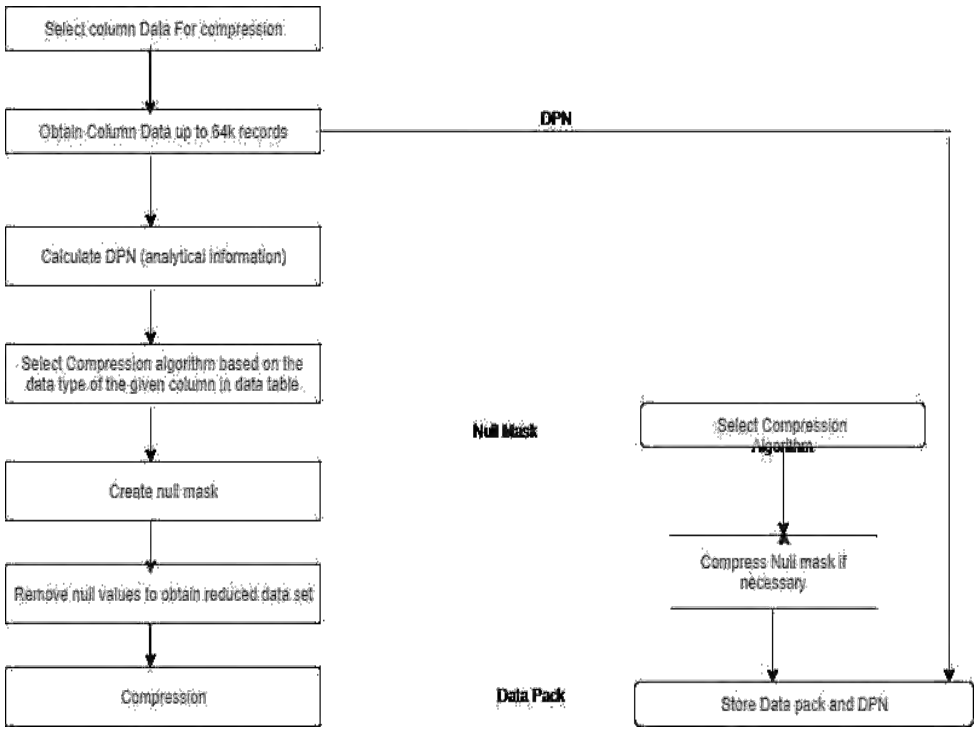


Figure 4. Data Storing and Compression

II. INFOBRIGHT COMMUNITY EDITION (ICE)

ICE is an open source software solution designed to deliver a scalable data warehouse optimized for analytic queries [4], [5].

A. *Key Benefits*

Ideal for data volumes up to 30TB

Market-leading data compression (from 10:1 to over 40:1), which drastically reduces I/O (improving query performance) and results in significantly less storage than alternative solutions

No licensing fees

Fast response times for complex analytic queries

Query and load performance remains constant as the size of the database grows

No requirement for specific schemas, e.g. Star schema

No requirement for materialized views, complex data partitioning strategies, or indexing

Simple to implement and manage, requiring little administration

Reduction in data warehouse capital and operational expenses by reducing the number of servers, the amount of storage needed and their associated maintenance costs, and a significant reduction in administrative costs

Runs on low cost, off-the-shelf hardware

Is compatible with major Business Intelligence tools such as Pentaho, JasperSoft, Cognos, Business Objects, and others.

B. *Infobright's Architecture*

It is based on the following concepts:

Column Orientation

Data Packs and Data Pack Nodes

Knowledge Nodes and the Knowledge Grid The Optimizer

Data Compression

The architecture is shown in Figure 2 .

1) Column Orientation

Infobright at its core is a highly compressed column-oriented data store. There are many advantages to column-orientation, including the ability to do more efficient data compression because each column stores a single data type (as opposed to rows that typically contain several data types), and allows compression to be optimized for each particular data type, significantly reducing disk I/O.

2) Data Organization and the Knowledge Grid

Infobright organizes the data into 3 layers:

a) Data Packs

The data itself within the columns is stored in 65,536 item groupings called Data Packs. The use of Data Packs improves data compression since they are smaller subsets of the column data (hence less variability) and the compression algorithm can be applied based on data type as shown in Figure 3.

b) Data Pack Nodes (DPNs)

Data Pack Nodes contain a set of statistics about the data that is stored and compressed in each of the Data Packs. There is always a 1 to 1 relationship between Data Packs and DPNs. DPN's always exist, so Infobright has some information about all the data in the database, unlike traditional databases where indexes are created for only a subset of columns.

c) Knowledge Nodes

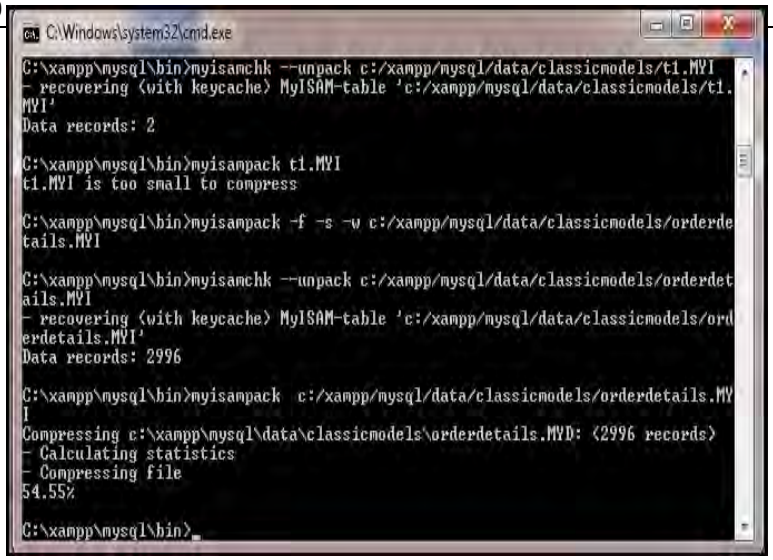


Figure 5. Compression command in MYSQL

These are a further set of metadata related to Data Packs or column relationships. They can be more introspective on the data, describing ranges of value occurrences, or can be retrospective, describing how they relate to other data in the database. Most KN's are created at load time, but others are created in response to queries in order to optimize performance.

3) The Infobright Optimizer

The Optimizer is the highest level of intelligence in the architecture. It uses the Knowledge Grid to determine the minimum set of Data Packs, which need to be decompressed in order to satisfy a given query in the fastest possible time.

4) Data Loading and Compression

The mechanism of creating and storing Data Packs and their DPNs is illustrated in Figure 4. One of Infobright's benefits is its industry-leading compression. Unlike traditional row-based data warehouses, data is stored by column, allowing compression algorithms to be finely tuned to the column data type. Moreover, for each column, the data is split into Data Packs with each storing up to 65,536 values of a given column. Infobright then applies a set of patent-pending compression algorithms that are optimized by automatically self-adjusting various parameters of the algorithm for each data pack. Within Infobright, the compression ratio may differ depending on data types and content. Additionally, some data may turn out to be more repetitive than others and hence compression ratios can be as high as 40:1. An average compression ratio of 10:1 is achieved in Infobright.

III. Implementation and Results

The PHP Hypertext Preprocessor (PHP) is a programming language that allows to create dynamic content that interacts with databases. This server-side scripting language designed for web development is used as a general-purpose programming language [6]. Since it provides easy connectivity to databases and its ease to understand it is chosen for implementation. XAMPP is used which provides Apache server to run the scripts and MYSQL which is a row oriented database [7].

A. Database used for Implementation

Once Infobright is downloaded and installed, you are ready to establish a connection to the Infobright columnar database. It's simply a matter of executing the following PHP code as you would with MySQL. In the rudimentary example we will use root for the user and 'password' as blank. [8].

B. Loading data in database:

Infobright includes a dedicated high performance loader that differs from the standard MySQL Loader. The Infobright Loader is designed for speed, but supports less LOAD syntax than the MySQL Loader, and only supports variable length text formatted load files [9].

a) Infobright Loader Syntax

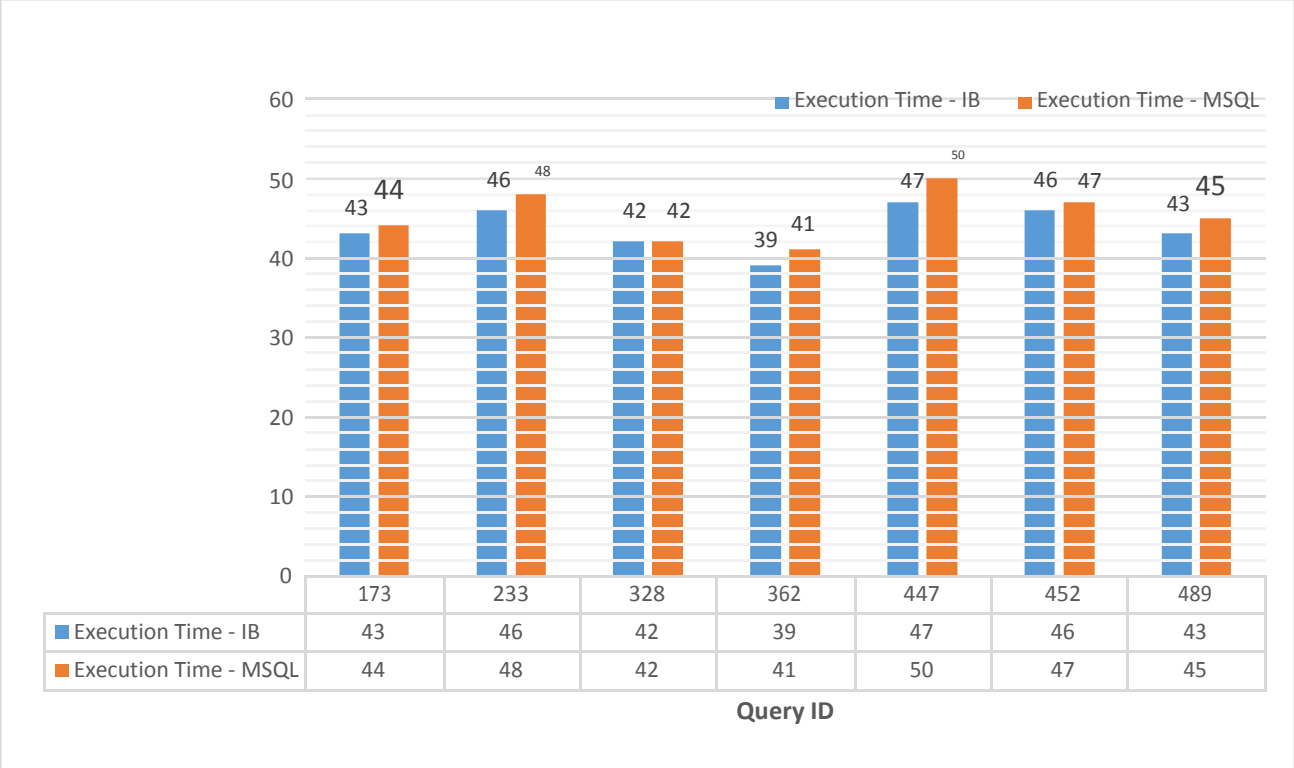
MY SQL Loader: Loads from a flat file that includes text and field delimiters. Supports more features for escaping embedded characters, error handling, and transforming data using functions at the time of load then the Infobright Loader. This loader is set using an environment variable (@bh_dataformat='mysql').

Import your data into an Infobright table by using the following load syntax (all other MySQL Loader syntax is not supported):

```
LOAD DATA INFILE
'/full_path/file_name' INTO TABLE
tbl_name [FIELDS
[TERMINATED BY 'char']
[ENCLOSED BY 'char']
[ESCAPED BY 'char']
];
```

Example: Loading data in OrderDetails Table in the database:

```
LOAD DATA LOCAL INFILE
'e:/datafiles/orderdetails.txt' INTO TABLE OrderDetails
```



	173	233	328	362	447	452	489
Execution Time - IB	43	46	42	39	47	46	43
Execution Time - MSQL	44	48	42	41	50	47	45

Figure 6. InfoBright and MySQL Query Execution Performance

FIELDS TERMINATED BY ',' ENCLOSED BY '"'
LINES TERMINATED BY '\r\n';
To load data in the database a form is provided to enter the name of the table.
Considering we have entered Orderdetails in the form above. In MYSQL the query executes in 0.211844 seconds. In Infobright the same query executes in 0.43 seconds. Note: The performance is affected in Infobright due to limitations of ICE (single threaded). In ICE due to multithreaded operation it is faster than MYSQL.

C. Compression:

One of the major advantages of a column-based approach is its effect on compression. Since data is stored by column, and each column consists of a single data type, it is possible to apply optimal compression algorithms for each column. This may seem like a small change but the difference in database size can be very significant when compared to other approaches. The compression information is stored in information schema database[10].
Compression in Infobright takes place automatically whereas in MYSQL we need to use pack command to compress and unpack to uncompress. Once compressed in mysql the data becomes readable only. Hence for every update we need to uncompress it.
Considering orderdetails table which has 2996 rows the compression is as shown in Fig. 6.

a) Syntax for mysql compression is:

Myisampack
c:/xampp/mysql/data/classicmodel/orderdetails.MYI

The compression is 54.55%. Hence the file size becomes 43.55KB. This can be shown from the information stored in information schema database. The query to show the size of the data file stored after compression is given by the query .
SELECT table_name AS "Table",round(((data_length+index_length)/1024), 2) "Size in KB" FROM information_schema. TABLES WHERE table_schema = "classicmodels" AND table_name = "orderdetails";

In Infobright compression takes place during loading and the compression is nearly 80%. Hence file size becomes 15.39 Kb. The same query above can be used to get the file size of the compressed file.

D. Data Analysis

Column oriented databases are better suited for analytics, where only portions of each record are required. By grouping the data together like this, the database only needs to retrieve columns that are relevant to the query, greatly reducing the overall I/O needed. Infobright is an example of an open source column oriented database built for high-speed reporting and analytical queries, especially against large volumes of data. How query is executed to retrieve data and display in graphical form.

Query showing the sum of all products ordered is given as:
SELECT sum(quantityordered) from orderdetails
Query execution time in MYSQL is 0.006655 seconds.

Whereas same query when executed in Infobright takes 0.002024 seconds.

Displaying Data Analyzed in Graphical Format

JpGraph is an Object Oriented graph drawing library for PHP 5.1 and above. Highlights of the available features are: multiple scale types; anti-aliasing of lines; color-gradient fills; support for GIF, JPEG, and PNG formats; and support for multiple Y axes. JpGraph can dynamically generate various types of graphs based on data you feed it from a database/array.

JpGraph makes it easy to draw both "quick and dirty" graphs with a minimum of code and complex professional graphs which requires a very fine grain control. JpGraph is equally well suited for both scientific and business type of graphs.

Query :

"SELECT sum(creditlimit),customernumber from customers group by customer number having sum(creditlimit) between 40000 and 50000"; In Infobright the execution time and bar chart for the same query is shown in Figure 6, InfoBright has less execution timing as compared toMySQL.

IV. CONCLUSION

The working of columnar databases for OLTP operations are better is presented in this paper on loading compression and data analysis speed. This is implemented using ICE as backend and PHP as frontend. The columnar databases for large volume of data are analyzed. Infobight database is having better performance 10% faster as compared to MySQL.

References

[1] Kanade A.S and Gopal A.,"Choosing right database system: Row or column-store", at IEEE International Conference on Information Communication and Embedded Systems, pp. 16-20, 21-22 Feb, 2013

[2] Gopal A. and Bhagat V., "Comparative study of row and column oriented databases", at Fifth International Conference on Emerging Trends in Engineering and Technology (ICETET), published in IEEE, pp.196-201, 5-7 Nov 2012.

[3] Infobright Community Edition 4.0.3 GA Quick Start by www.infobright.org

[4] Infobright Analytic Database: Architecture, 2012 by www.infobright.org

[5] Download software at www.infobright.org/download

[6] Dalibor D. Dvorski, "Installing, configuring, and developing with xampp", March 2007

[7] XAMPP at https://www.apachefriends. org/download.html

[8] Infobright, "DataLoadingGuide: Revision2.1", November11,2010

[9] Databasefromwww.eclipse.org/birt/documentation/sample-database.php

[10] "Upgrading Your Row-based Database to a Columnar Database in Infobright", 2010.

Researcher Framework using MongoDB & FCM clustering for Prediction of the Future of Patients from EHR

Harish Barapatr,Dr. Yogesh Sharma,Julli Sarode,Vaishali Shinde4
Dept. Of Computer Engineering
Yadavrao Tasgaonkar Institute Of Engineering And Technology
B.R. Harne College OF Engineering, Vagani.
harishkbarapatre@gmail.com,julli.sarode01@gmail.com,vaishalishinde22@gmail.com

Abstract

Data collection and data analysis models changed business trends over the past few years. By using BigData analytics we can predict the effects of drugs and how drugs develop disease on mankind. Many machine learning algorithms like Cluster computing environment, Classification etc. analyze the content of Health care. The proposed framework developed C-means clustering algorithm in developing Biomedical Engineering applications. The data is collected from machine learning repository. Big data framework dbmongo database is used to analyze the data. Here we have modules like Doctor, Administrator and analyst. By using machine learning and Bigdata, the analyst module identifies the chronic disease and help in reducing the medical expenses. Doctor by observing the predicted data symptoms and predicts related disease. Administrator's role is to add or remove the users in the database.

Keywords: mongoDB, BigData, C-means clustering, K-means clustering, Supervised Learning

I.INTRODUCTION

BigData is widely used in Biomedical engineering domain in the recent trends. The Research studies have been carried out in BigData analysis so, as to work on the domains like BioMedical Engineering. Before discussing all this it has to be discussed about the reason we are choosing BigData [1][2] [3] [4].

This paper highlights supervised learning techniques where there is a need of training data that is acquired with training examples. Supervised learning is a machine learning techniques which always maps the input with the output basing on the examples that were provided to it. Supervised learning always works with supervisor or trainer that we provide to it. Supervised learning builds model and later it evaluates it with predictions as base. It trains the input data with the given examples using the stated algorithm and make reasonable predictions. So, as to solve problem of supervised learning, the following steps are performed:

Step1. Initially we have to decide the type of training examples.

Step2.Secondly there is need for collecting a training set.

Step3: Determine the learned function for the input representation.

Step4 Learned function structure is determined with corresponding learning algorithm.

The above mentioned research paper Researcher Framework using MongoDB & FCM clustering for Prediction of the Future of Patients from EHR is said to help the patients .

II. OBJECTIVES

The main objective of this masters degree project

1. Is to examine different clustering algorithms in order to detect groups in a real-world, high-dimensional EHR data set.
2. The goal of the work is to find algorithms that detect high-quality clusters Accuracy within a reasonable computational time.
3. In order for the chosen algorithms to scale with an increasing size of the patient sample size and/or features, it is desirable.
4. To find an algorithm that achieves high-qualityAccuracywithminimal preliminary processing of the data

The choice of methods for the identification of groups in the data set imposes challenges due to several aspects:

• High-dimensionality: The data set originally has about 1200 features, which constrains the choice of algorithms and evaluation measures.

• Missing data: The set contains missing values that are not missing at random.

• Accuracy: Data are collected through health questionnaires and the accuracy of this self-reported data is unknown

III. EXISTING SYSTEM

Indeed hierarchical clustering algorithms generally seems to perform better than partitioned methods in creating high-quality clusters. As opposed to partitioned methods, they are also able to discover non spherical clusters. Hierarchical clustering algorithms also offer a good visualization in the form of a dendogram, which is suitable for interpretation. The hierarchical methods has the main drawback that the time complexity is quadratic O(n2).Because of the popularity in literature and widespread use of the two traditional methods, hierarchical clustering and C-means clustering will be of interest in this study. Although our high-dimensional data set might implicate lengthy running times for the hierarchical method, leading to lengthy experiments, it compensates this drawback through producing high-quality clusters, which is of main interest for this work.

IV. PROBLEM DEFINITION

Effective prediction of clinical risks of ACS patients via their heterogeneous EHR data is still an intricate problem and remains a major challenge for health care management, mainly due to high clinical complexity and the natural heterogeneity in EHR data. Therefore, one of the most important tasks in clinical risk prediction is to develop robust prediction models that can effectively handle high dimensional heterogeneous EHR data and accurately classify different clinical risks levels based on the acquired EHR data. It has particular advantages such as rapid inference and the ability to reconstruct features (a.k.a. clinical risk factors) yielding good classification accuracy. In order to ensure that features reconstructed by FCM are useful to the HER in problem for each patient, two regularization constraints are added on FCM to capture characteristics of patients at similar risk levels to make the reconstructed features of patients within the same risk level as close as possible, and preserve the discriminating information across different risk levels to make the reconstructed features of patients as separated as possible. After this we append FCM Clustering layer on the top of the resulting reconstructed feature representation layer, which is tailored to the EHR problem. Our proposed model learns more discriminative patient feature representations and thus improves the performance of EHR Data set.

V. ROLE OF BIGDATA AND MACHINE LEARNING IN BIOMEDICAL ENGINEERING

BigData and Machine Learning helped BioMedical Engineering in many aspects [1][2][3][4][5][6][7].

[1].Artificial Intelligence came into existence from the study pattern recognition and computational learning theory.

[2].Machine learning creates algorithms that learn from many examples and it makes predictions on the data

[3].Machine learning algorithms were trained to Identify the disease basing on symptoms. It helps to Diagnose Parkinson's Disease based on Dysphonia Measures. It also provide a review of recent researches on brain tumor etc

[4].Machine-learning techniques used for identifying Retinal image analysis.

[5].Machine learning algorithms supported in finding Obstructive Sleep Apnoea Detection using neural network with the help of deep learning framework.

[6].Machine learning algorithms help in detecting cell damage and tumors.

[7].Researches not only detects the disease they also try to provide remedies to the concerned disease.

VI. PRAPOSED SYSTEM

The proposed framework contains modules like Administrator, Doctor(Clinician) and Analyst / Researcher. The Administrator role is to add or remove users i.e. update user information. The Doctor's module role is to name the disease and their symptoms. All the disease regarded information is stored in database. Analyst module role is to provide the parameter values for the analysis and it has to apply relevant C-means algorithm to the provided data. The parameters can be like names, dates, gender or age.

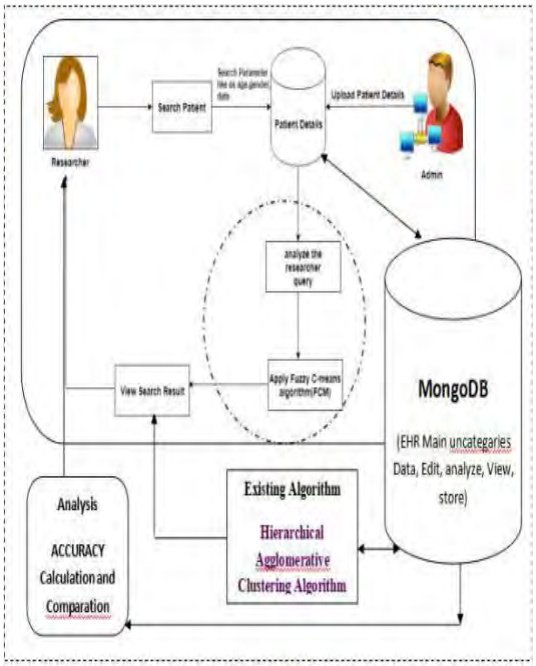


Figure No: 1- SYSTEM ARCHITECTURE

Figure No-1 shows the framework flow which explain the working of the system.

Analyst after choosing required parameter, he/she can select the required representation method to get the desired output. Proposed Frame Work where the Bigdata trained with machine learning using Fuzzy. C-Means algorithm. Hierarchical clustering iteratively joins the two nearest clusters beginning from singleton clusters or with the complete set. After joining of two clusters distance between all the other clusters and a new cluster that is joined are recalculated.

a) PATIENT DATA

The hospital dataset used in this study contains real-life hospital data, and the data are stored in the data center. To protect the patient's privacy and security, we created a security access mechanism. The data provided by the hospital include EHR, medical image data and gene data. The inpatient department data is mainly composed of structured and unstructured text data. The structured data includes laboratory data and the patient's basic information such as the patient's age, gender and date, etc. While the unstructured text data includes the patient's narration of his/her illness, the doctor's interrogation records and diagnosis, etc. In order to give out the main disease which affects this region, we have made a statistics on the number of patients, the sex ratio of patients and the major disease in this region every year from the structured and unstructured text data.

b) PATIENT DETAILS PREDICTION

We obtain the main cancer disease in this region. The goal of this study is to predict whether a patient is amongst the cerebral infarction high-risk population according to their medical history. More formally, we regard the risk prediction model for cerebral infarction as the supervised learning methods of machine learning, i.e., the input value is the attribute value of the patient, A = (a1, a2, · , an) which includes the patient's personal information such as age, gender, the prevalence of symptoms, and other structured data and unstructured data.

c) ALGORITHM ACCURACY

The accuracy determines and predicts the measure of how machine learning classification algorithm is working correctly. It finds the number of true positives and true negatives and is also used to divide by the number of true positives and negatives, false positives and negatives. True positive /true negative is a data which determine that algorithm is designed. False positive / False negative determine that algorithm is incorrect it is not classified properly. It we get correct result it is true positive.

d) MACHINE LEARNING

Machine learning (ML) algorithms help software applications to predict outcomes accurately. Machine learning follows the same strategies of data mining and predictive modeling. Machine learning is widely used in many domains like fraud detection, network security threat detection etc.

Fuzzy C-means clustering algorithms steps are:

1. Intially Choose C cluster centers in a random. (matrix V).
2. Make Computation for the membership matrix U.
3. Make modification by Updating the cluster Centers vi.
4. Objective function J is to be computed.
5. We have to repeat steps 2 to 4 until we get Convergence (|| var (J) ≤ €||).

6.After execution of all steps are completed w we Exit.

VII. SOFTWARE DETAILS

In this section, drugs for the patients are chosen basing on K-means clustering. We used software like [1][2][3][4].

[1] Operating System used for the research is Windows 8.

[2] Application Server used is Tomcat 6.0/8.0

.[3] Front End used is HTML, Java,Jsp.

[4] Script used is JavaScript. Database used is MongoDB. Database Connectivity used is Robomongo-0.8.5-i386.

Why we are Using MongoDB?

1. The storage pattern in MongoDB is Document Oriented Storage. Here the Data is stored in the form of JSON(JavaScript Object Notation (JSON) pattern style documents.

2. MongoDB uses multikey indexes . The indexing of multikey index is used to store the content in arrays. Field is indexed that holds an array value.

3. MongoDB creates variable index entries for each and every element of the array. Multikey indexes allow queries to select the documents which contain arrays by matching .A replica set in MongoDB is a group of mongodb processes that can maintain the same data set.

Replica sets in MongoDB provides redundancy and high availability.

4. Sharding feature of MongoDB is the process of distributing data across multiple servers for storage purpose. MongoDB software uses sharding to manage huge massive data growth. Sharding feature in MongoDB allows to add many servers to our database to support the data growth. MongoDB automatically balances data and load across its various servers.

Advantages of MongoDB over RDBMS

MongoDB database is a document based schema less in which database where there is collection that holds variety of documents. The Number of fields, the content and size of document differ from one another.

Structure of object in MongoDB is very clear.

MongoDB database have no complex joins to work with.

MongoDB supports dynamic queries when we are working with documents. This is called document-based query language which is as powerful as that of Structured Query Language.

MongoDB is much easier to scale.

No need of Object to database mapping in MongoDB.

MongoDB uses internal memory for storing the working setand that help in fast access of data.

VIII. ALGORITHM

Fuzzy C means algorithm

In cluster populations or data points are divided into different groups. Data points of same group are similar and data points of dissimilar group are not similar. C-means clustering algorithm works by assigning membership to each data point. The member ship to data point is given basing on data point distance with cluster centre. If the data point is nearest to cluster centre its membership is more. We can conclude the data point nearest to cluster centre serve as prototype of cluster

Fuzzy C-means clustering algorithms steps are:

1. Chose C cluster centers randomly (matrix V)
2. Compute the membership matrix U
3. Update the cluster centers vi
4. Compute the objective function J
5. Repeat steps 2 to 4 until convergence (|| var (J) ≤ €||)
6. Exit.

Data Base with MongoDB

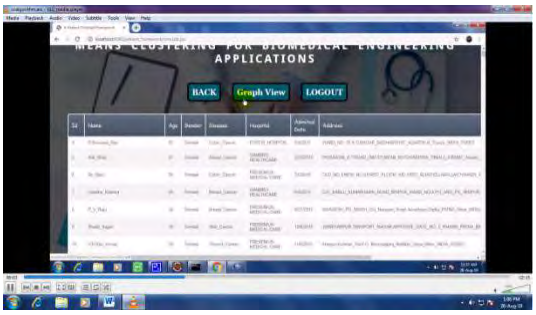


Figure No-2 : Mongodb Scinerio-1

Figure No-3 : Mongodb Scinerio-2

IX. RESULT ANALYSIS

1) This C-means algorithm gives best results for overlapped datasets when compared with k-means algorithm.

2) In k-means algorithm data points are assigned to single cluster center. In c-means algorithm data points are assigned its membership to more than one cluster

Accuracy of C-means clustering algorithm

Calculation Formula

Accuracy can be calculated as below

Acc = (Pc / Ds)

(Accuracy=predicted class/dataset)

Acc – Stands for accuracy

Pc - Stands for correctly predicted

class Ds - Stands for Dataset

NM-Not match User variable

MMngMU

Where $X_i=X_1+X_2+X_3\ldots+X_n$

Where xi represents a measurement and n is the number of measurements.

Modified Fuzzy C-Means Clustering

$$J(U,V)=\sum_{i=1}^{c}\sum_{j=1}^{N}U\binom{m}{ij}||x_{j-}v_i||^2$$

Where

- U: membership matrix
- V: cluster centers matrix
- N: number of data points
- C: number of clusters
- Xj: is the jth measured data point
- vi: is the center of cluster i

Dunn was the first to implement fuzzy C-means clustering algorithm in 1973, later Bezdek improved this algorithm in 1981. For fuzzy C-means, the membership function is represented by matrix (U) having values between 0 and 1 that represent the membership of data points to every cluster, while hard C-means uses only 0 and 1 values for membership function. Fuzzy C-means algorithm is based on an objective function.

The search result of gender parameter using FCM are shown below

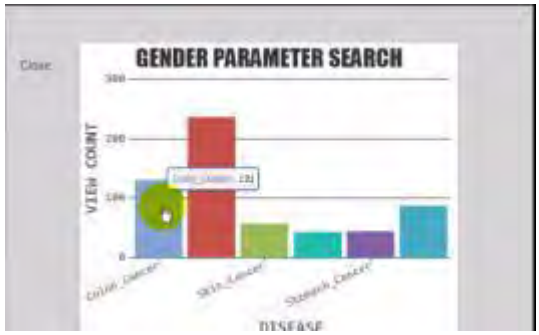


Figure No-4 : analysis with gender parameter

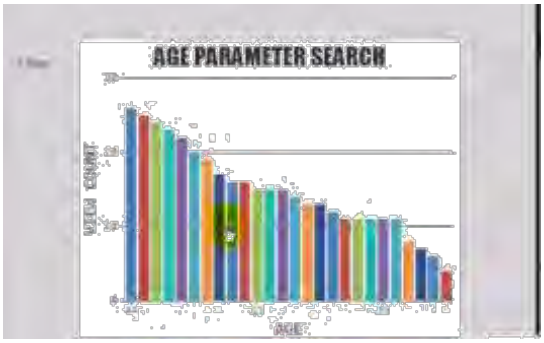


Figure No-5 : Analysis with age parameter

X. CONCLUSION

The aim of this paper was to propose a framework using Bigdata and K-means clustering method to select drugs for patients. Biomedical engineering applications are trained with k-means clustering algorithm. This is a deep learning approach. Machine learning and BigData Analytics together helps in developing a new era in Biomedical engineering. From Identifying Diseases and Diagnosis to Machine Crowd sourced Data Collection the machine learning plays a major role. Both Bigdata science and the approach of machine learning together build Biomedical engineering domains. They are dependent upon one another and they are mutuality. Bigdata analytics makes analysis on huge volumes of data and machine learning helps to give input and get output patterns. Hadoop and Mapreduce are used for distributed processing of large datasets. MongoDB is a NoSQL database used for high volume data storage and it is document oriented. The database used here is MongoDB. Here in this research C-means algorithm is taken for training purpose as it works faster. C-means algorithm assigns its values to multiple clusters. In the proposed framework the work has been done on C -means clustering algorithm. So, selecting C-means made the programming part simple. In future this above said framework will be developing
Healthcare data solutions, Healthcare Intelligence and data accuracy etc.

REFERENCES

[1] . F. F. Costa, “Big data in biomedicine,” Drug discovery today, vol. 19, no. 4, pp. 433–440, 2014.

[2] C. Auffray, R. Balling, I. Barroso, L. Bencze, M. Benson, J. Bergeron,E. Bernal-Delgado, N. Blomberg, C. Bock, A. Conesa et al., “Making sense of big data in health research: towards an eu action plan,” Genome medicine, vol. 8, no. 1, p. 71, 2016.

[3] Q. Yao, Y. Tian, P.-F. Li, L.-L. Tian, Y.-M. Qian, and J.-S. Li, “Design and development of a medical big data processing system based on hadoop,” Journal of medical systems, vol. 39, no. 3, p. 23, 2015.

[4] Tehrani, A.; Lee, H.; Mathews, S.; Shore, A.; Makary, M.; Pronovost, P. Diagnostic Errors More Common, Costly and Harmful Than Treatment Mistakes; John Hopkins Medicine: Baltimore, Maryland, 2013.

[5] Ongenae, F.; Claeys, M.; Dupont, T.; Kerckhove, W.; Verhoeve, P Dhaene, De Turck, F. A probabilistic ontology-based platformfor
self-learning context-aware healthcare applications. Expert Syst. 2013, 40, 7629–7646.

[6] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare

communities,” IEEEAccess, vol. 5, pp.

8869–8879, 2017.

[7] R. Mao, H. Xu, W. Wu, J. Li, Y. Li, and M. Lu, “Overcoming the challenge of variety: big data abstraction, the next evolution of data management for aal communication systems,” IEEE Communications Magazine, vol. 53, no. 1, pp. 42–47, 2015.

[8] B. M. Eskofier, S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz et al., “Recent machine learning advancements in sensor-based mobility analysis:
deep learning for parkinson’s disease assessment,” in Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE, 2016, pp. 655–658.

[9] N. K. Mishra and M. E. Celebi, “An overview of melanoma detection in dermoscopy images using image processing and machine learning,” arXiv preprint arXiv:1601.07843, 2016.

[10] Sajid I, Khan UG, Saba T, Rehman A.

Computer-assistedbraintumortype
discrimination using magnetic resonance

imaging features. Biomed Eng Lett. 2018.

[9] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”.In: Journal of Machine Learning Research 12 (2011), pp. 2825–2830.

[10] Deborah Lupton and Annemarie Jutel. “’It’s like having a physician in your pocket!’ A critical analysis of self-diagnosis smartphone apps”. In: Social Science and Medicine 133 (2015), pp. 128–135.

[11] Michael E. Houle, Hans Peter Kriegel, and Peer Kröger. “Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?” In: Proceedings of SSDBM. 2010, pp. 482–500.

[12] Nenad Tomašev and Miloš Radovanovi´c. “Clustering Evaluation in High-Dimensional

Data”. In: Unsupervised Learning

Algorithms.Springer, Cham, 2016, pp. 71–107.

A Review of Deep Learning Techniques used in Breast Cancer Image Classification

Dr Vidya Sarode ,Dr Amog Chaudhari,Fabian T. R. Barreto
Department of Electronics & Telecom. MBBS,
Xavier Institute of Engineering MGM Institute of Health Sciences
Mumbai, India v
Vidya.s@xavier.ac.in,amogh.chaudhari@gmail.com,frfabiansj@xavier.ac.in

Abstract—**Human society is faced with the increasing global burden of cancer as we enter the new decade. According to the Globocan 2018 report, in India alone, breast cancer has the largest number of new cases (27.7 %) as compared to the world incidence rate of 24.7 %. Breast cancer detection after clinical examination includes two tests, namely Breast Imaging and Breast Histopathology. Breast cancer has different stages or spread, aggressiveness, and genetic makeup. An end to end system that helps in early detection and prevention would increase the survival rates. Traditionally, Machine Learning techniques have been used to detect malignancy in breast cancer images. However Machine Learning is limited in its ability to process natural data in its raw form. The limitation is due to the need of domain expert who can carefully handcraft features to feed a classifier. Deep Learning, a subfield of Machine Learning, however automatically learns multiple levels of representation and abstraction that help to understand breast cancer images at a greater depth. Deep Convolutional Neural Networks have achieved remarkable breakthroughs in processing images. In our paper we review the Deep Learning techniques that have given extraordinary performance for both Breast Cancer Image and Breast Cancer Histopathology Image Classification and thus helped in early detection of Breast Cancer.**

Index Terms—Breast Cancer Image Classification, Convolutional Neural Networks, Deep Learning, Machine Learning

I. INTRODUCTION

As we begin a new decade, we are faced with the increasing global burden of cancer. It remains an unresolved cause for increased morbidity and mortality of human beings. This in turn impacts societal life and economy. The GLOBO-CAN database, compiled by the International Agency for Research on Cancer (IARC), provides estimates on national cancer incidence and mortality [1]. In India, according to the GLOBOCAN 2018 report, breast cancer has the largest number of new cases (27.7 %) as against the world incidence rate of 24.7 %.

The last decade has seen an intensified research efforts in the field of prevention and early detection of this disease. As the wave of rapid urbanization sweeps India, one observes drastic changes in dietary habits and life style, causing at times high stress levels. Studies in high-income countries have indicated that from one-third to two-fifths of new cancer cases could be avoided by eliminating or reducing exposure to known lifestyle and environmental risk factors [2].

The risk factor for most cases of breast cancer is due to a combination of genetic and environmental factors. The breast cancer detection process, after physical examination includes two tests, namely Breast Imaging and Histopathology. Early detection, true diagnosis and prompt treatment are crucial for a better prognosis [3]. This is made possible by effective and efficient screening programs. Early diagnosis of cancer in the natural progression of disease, before it spreads to local regions or to distant organs can result in higher 5-year survival rates.

Imaging modalities may be classified as primary or simple and secondary or complex. In regular screening of patients and as a first line of diagnostic tools only the primary imaging modalities are preferred because they are cheap and provide reliable and repeatable results under a variety of conditions. One of the commonly used imaging modalities for regular screening of patients for breast cancer is mammography. Here any anomaly is categorized as regular or irregular, and capsulated (benign - not cancerous) or non-capsulated (malignant). Mammography techniques may be a screen-film mammography (SFM) or a full-field digital mammography (FFDM). SFM may not detect benign cancer properly. FFDM performs better in case of misdiagnosed cancer samples as the images are processed using Image processing techniques.

Screening mammography reduces breast cancer mortality by about 20% to 35% in women aged 50 to 69 years and slightly less in women aged 40 to 49 years [4]. Some of the other primary imaging modalities include Ultrasound, Thermogra-phy and Electrical Impedance. For further examination and confirmation of dier screening has potential limitations and drawbacks which include false-positive, false-negative test results and overdiagnosis.

Presently, full-field digital mammography (FFDM) is rapidly being replaced by digital breast tomosynthesis (DBT, also called 3D mammography). An independent risk factor for breast cancer is the breast tissue density. It results in reduced mammographic accuracy because of masking and superimposition of dense tissue. Hence the need for supplemental imaging to improve the sensitivity of screening in women with significant mammography breast density [5].

Mammographic breast density is visually assessed by radiologists in routine mammogram image reading, using four qualitative Breast Imaging and Reporting Data System (BI-RADS) breast density categories. However it is difficult for radiologists to consistently distinguish the two most common and most variably assigned BIRADS categories, i.e., "scattered density" and "heterogeneously dense" [6].

II. MACHINE LEARNING AND DEEP LEARNING

A. Machine Learning

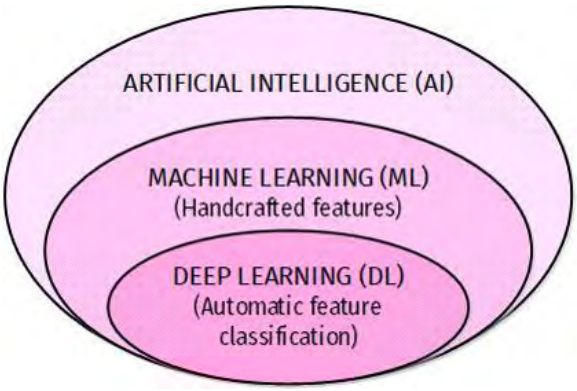


Fig. 1. Difference between ML and DL

The pioneer of Machine Learning (ML), Arthur Samuel, defined ML as a "field of study that gives computers the ability to learn without being explicitly programmed." ML primarily focuses on classification and uses regression analysis on known features previously learned from the training data. ML helps to understand the structures and patterns within large datasets. It is used for classification, clustering and prediction. Classification is putting something into one or more categories. Clustering is dividing a large set of data points into a few clusters, with points within a cluster having common properties. Prediction uses historical data and builds models that are used to forecast a future value.

Conventional ML techniques are limited in their ability to process natural data in their raw form. The limitation is because of the need of domain expert and carefully hand feature extraction method as shown in Figure 1. The low-level features can be hand crafted with great success for certain data and tasks, but designing effective features for new data and tasks usually requires new domain knowledge as most hand-crafted features cannot simply be adapted to new conditions [7]. ML algorithms uses Support Vector Machines (SVMs) with different kernel functions like Gaussian Radial Basis Function (RBF) or a polynomial function. SVMs are "non-parametric" models where the parameters are not predefined and their number depends on the training data used. An SVM classifier performs binary classification. Multiple classifiers may be combined using the bagging and boosting techniques to improve performance.

The other ML techniques that have been used are k-Nearest Neighbors (k-NN), Random Forest (RF) and Bayesian Networks (BNs) [8]. k-NN captures the idea of similarity (distance or closeness) and votes for the most frequent label in case of classification. RF combines many decision trees to ensemble a forest of trees. This gives a better stability and is also insensitive to the noise of the input data. BN is a subfield of probabilistic graphical models (directed acyclic graph - DAG) that are used for prediction and knowledge representation of uncertain domains. In ML the classification is done in two stages, namely, training and testing. During training the image is acquired, preprocessed and features are extracted and selected. Similar steps are followed during testing. The preprocessing step helps to remove the noise and thus improve the quality of the image. Finally in the classification stage different classifiers are applied for diagnosis and prognosis.

B. Deep Learning

Deep Learning (DL) uses Deep Neural Network (DNN) architectures as shown in Figure 2, to learn multiple lev-els of representation and abstraction. DL techniques are representation-learning methods that help to understand data such as text, images and sound. The features are learned automatically by the DNN and are obtained by composing simple but non-linear modules that each transform the representation at one level (beginning with the raw input) into a representation at a higher, slightly more abstract level. These transformations can help to learn very complex intricate functions. A deep-learning architecture is thus a multilayer stack of simple modules, which learn features and many of which compute non-linear input–output mappings.

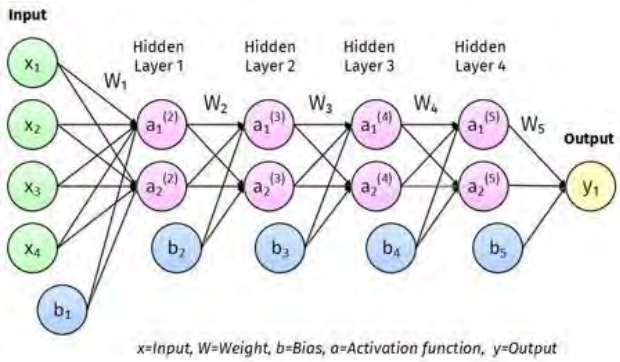


Fig. 2. Deep Neural Network (DNN) architecture

The power of DNNs is due to the highly optimized hardware, that is the Graphics Processing Units (GPUs) which can handle lots of parallel computations using thousands of cores and possess a large memory bandwidth to deal with the data for these large computations. They work on the massive training data and learn its multi-level feature representation. Li Deng [9] categorizes DL architectures into thre

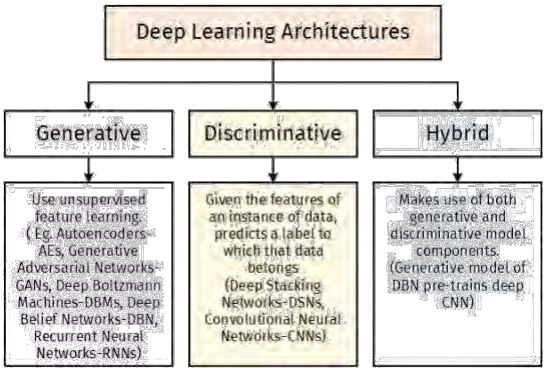


Fig. 3. Deep Learning Architectures

III. DATASETS

In the past three decades several image databases have been created for breast cancer research, some of which are public and others private. Public databases provide a common reference for researchers to test and compare their methods. The image database is crucial to the type of study that one wishes to undertake. Often the collected datasets are class imbalanced. There is a large class where the patients do not have cancer and a small class of patients with cancer. Necessary precautions need to be taken while conducting research by choosing the proper research design.

Datasets are available both for Breast Imaging and Breast Histopathology. Based on the imaging results, a doctor may suspect that a person has breast cancer. A biopsy is then performed to diagnose if cancer is really present. Among the different biopsy techniques, the most common ones are fine needle aspiration (FNA), core needle biopsy, vacuum-assisted, and incisional or excisional surgical biopsy [10]. The surgical biopsy, though reliable, is invasive and costly.

The designed classifying system needs an adequately large and preferably public database. In case of Deep learning this database has to be large to get meaningful classification results. The author in [11] give the details of the requirements for mammographic databases. These include the case selection, ground truth, patient's associated information and importantly the organization and distribution of the database. In [12] a case is made for Computer Aided Detection (CADe) as against a human person interpreting screening mammograms because it is a highly repetitive task and also highly subjective, resulting in large intra- and inter-interpreter variability.

Breast Cancer Wisconsin (Diagnostic) Data Set [13] which was made public in 1992, has been widely used especially for creating good machine learning classifiers [14]. It has 32 attributes, 569 instances and two classes. Another database is the mini-MIAS (Mammographic Image Analysis Society) database of mammograms [15]. The dataset contains 322 images, with each image falling into one of seven categories: calcification, circumscribed masses, spiculated masses, architectural distortion, asymmetry, other ill-defined masses and normal. The size of each image is 1024x1024 pixels.

A publicly available database, is the Digital Database for Screening Mammography (DDSM) [16] which is available from the University of South Florida was released in 1997. The authors in [17] have developed a new dataset called INbreast. It has total of 115 cases (410 images) from which 90 cases are from women with both breasts affected (four images per case) and 25 cases are from mastectomy patients (two images per case).

The Breast Cancer Histopathological Image Classification (BreakHis) is composed of 9,109 microscopic images of breast tumor tissue collected the through a clinical study from Jan-uary 2014 to December 2014. It is collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). It contains 2,480 benign and 5,429 malignant samples which are 700X460 pixels, 3-channel RGB, 8-bit depth in each channel in the PNG format [18].

IV. DEEP LEARNING MODELS

AlexNet [19] is a deep convolutional neural network that classified 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. The neu-ral network has 60 million parameters and 650,000 neurons. It consists of five convolutional layers, two of which are followed by max-pooling layers, and at the end are three fully-connected layers with a final softmax layer as shown in Figure 4.

The authors in [6] used a improved version of the AlexNet model. They used the two-class CNN model to classify the two BI-RADS breast density categories: "scattered density" vs. "heterogeneously dense". They have suggested future work in the area of identifying more likely to be misclassified images, comparing other CNNs model structures to the AlexNet, developing other approaches to deal with noisy or potentially inaccurately labeled data and testing by using larger multi-center datasets. They also used transfer learning from nonmed-ical images to medical imaging-based applications. They used the AlexNet model pretrained on ImageNet and fine-tuned it with their mammogram data.

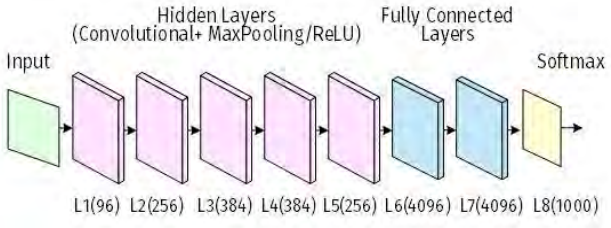


Fig. 4. AlexNet Architecture

concatenations). The authors in [21] have developed a deep learning model called Grouped-Resaunet (GRA U-Net) which

U-Net [20] describes a Convolutional Neural Network (CNN) that was developed for biomedical image segmenta-tion at the Computer Science Department of the University of Freiburg, Germany. The architecture works with fewer training images and yields more precise segmentations. The name comes from the U-shaped architecture consisting of a contracting (Convolution + Maxpooling + Rectified Linear Unit-ReLU) and an expanding path (up-convolutions and

segments the nipple region with very high accuracy from other similar looking non-nipple regions such as tumour.

The authors in [22] have developed an efficient deep learning framework. It identifies, segments, and classifies cell membranes and nuclei from human epidermal growth factor receptor-2 (HER2)-stained breast cancer images with minimal user intervention. The also used Long Short-Term Memory (LSTM) so as to use pixel information from the multiple frames in making the semantic segmentation and classification cases. Their model achieved 96.64% precision, 96.79% recall, 96.71% F-score, 93.08% negative predictive value, 98.33% accuracy, and a 6.84% falsepositive rate. Their framework was run for almost 7 days (24 7 = 168 hours) for training that was done on a parallel computing platform. It took on average, 1.06 seconds for cell membrane and nucleus segmentation, classification and HER2 scoring on test data.

V. CONCLUSION

As researchers fight against time to discover the holy grail to cure cancer one has to accept the fact that great progress has been achieved to increase the survival rate. Deep Learning-based CNN models have given extraordinary performance for both Breast Cancer Image and Breast Cancer Histopathology Image Classification. As Deep Learning progresses, we look forward to an end to end system, which along with other mul-tidisciplinary approaches, will not only help in early detection but also help prevent cancer in the first place.

REFERENCES

[1] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. M. Parkin, M. Pineros,˜ A. Znaor and F. Bray, "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," International journal of cancer, vol. 144, pp. 1941-1953, 2019.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: a cancer journal for clinicians, vol. 68, pp. 394-424, 2018.

[3] "Early Detection of Breast Cancer," [Online]. Available: http://www.breastcancerindia.net/screening/guidelines.html. [Accessed 10 January 2020].

[4] J. G. Elmore, K. Armstrong, C. D. Lehman and S. W. Fletcher, "Screening for breast cancer," Jama, vol. 293, pp. 1245-1256, 2005.

[5] R. C. Wender, O. W. Brawley, S. A. Fedewa, T. Gansler and R. A. Smith, "A blueprint for cancer screening and early detection: Advancing screening's contribution to cancer control," CA: a cancer journal for clinicians, vol. 69, pp. 50-79, 2019.

[6] A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz and S. Wu, "A deep learning method for classifying mammographic breast density categories," Medical physics, vol. 45, pp. 314-321, 2018.

[7] Y. Bengio, A. Courville and P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, pp. 1798-1828, 2013.

[8] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," in 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), 2016.

[9] L. Deng, "Three Classes of Deep Learning Architectures and Their Applications: A Tutorial Survey," APSIPA Transactions on Signal and Information Processing, 1 2012.

[10] "Biopsy," 7 October 2015. [Online]. Available: https://www.breastcancer.org/symptoms/testing/types/biopsy. [Accessed 12 January 2020].

[11] R. M. Nishikawa, "Development of a Common Database for Digital Mammography Research.," Chicago, IL: University of Chicago, 1996.

[12] R. M. Nishikawa and D. Gur, "CADe for early detection of breast cancer—current status and why we need to continue to explore new approaches," Academic radiology, vol. 21, pp. 1320-1321, 2014.

[13] "Breast Cancer Wisconsin (Diagnostic) Data Set," [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic). [Accessed 9 January 2020].

[14] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," Optimization methods and software, vol. 1, pp. 23-34, 1992.

[15] "The mini-MIAS database of mammograms," 11 December 2012. [Online]. Available: http://peipa.essex.ac.uk/info/mias.html. [Accessed 8 January 2020].

[16] "University of South Florida Digital Mammography," [Online]. Available: http://www.eng.usf.edu/cvprg/Mammography/Database.html. [Accessed 9 January 2020].

[17] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," Academic radiology, vol. 19, pp. 236-248, 2012.

[18] "Breast Cancer Histopathological Database (BreakHis)," [Online]. Avail-able: http://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/. [Accessed 10 January 2020].

[19] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012.

[20] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, 2015.

[21] Z. Zhuang, A. N. J. Raj, A. Jain, N. Ruban, S. Chaurasia, N. Li, M. Lakshmanan and M. Murugappan, "Nipple Segmentation and Localization Using Modified U-Net on Breast Ultrasound Images," Journal of Medical Imaging and Health Informatics, vol. 9, pp. 1827-1837, 2019.

[22] M. Saha and C. Chakraborty, "Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation," IEEE Transactions on Image Processing, vol. 27, pp. 2189-2200, 2018.

Clinical Risk Prediction of Acute Coronary Syndrome Using Deep Learning Approach

Ms. Swati Suchak, Dr. Uttara Gogate
PG Scholar, Associate Professor
Alamuri Ratnamala Institute Of Engineering & Technology, Shivajirao S. Jondhale College Of Engineering,
Mumbai, India,
swatisuchak0702@gmail.com,uttara.gogate16@Gmail.Com

Abstract: **Acute coronary syndrome (ACS) is a severe cardiovascular disease which is leading to death and it is serious long-term disability globally. Prediction of ACS is important for diagnosing earlier to treat it. Previous ACS models based on small set of risk factors and predictive variables used to simplify the score calculation are numbered. This study developed stacked and regularized un noising auto-encoder (SDAE) model to find clinical risks of ACS patients from huge collections of electronic health records (EHR). It Determine patients at similar risk levels characteristics, low risk, high risk and medium risk and preserve the results. This approach is validated with more than 2000 real clinical dataset consisting of patient samples. The Approach followed by us remains robust and gives more than 0.8 accurate results for predicting AUC.**

Index Terms—Acute coronary syndrome, Clinical risk Deep learning, Electronic Health Records

I. INTRODUCTION

Acute coronary syndrome (ACS) includes unstable angina and myocardial infarction (MI) are the common types of coronary heart diseases (CHD).This disease occurs if the heart muscle does not receive sufficient oxygen in blood. Due to this patient may face problems like morality, ACS became an important health issue its treatment cost and the prolonged chronic course mortality rates increased.[4]. Every year ACS is estimated to affect millions of people in china and U.S.A.[5]. Approximately 1/3 rd population are at risk of ACS disease.[6] In United States of America more than50, 000 to 60,000 people are dying before taking them to hospital because of ACS attack. All the above stated adverse affects have to be ended by diagnosing patients and prevent this disease.[7] Prevention of ACS may help both the patients and society. These are mentioned in points. [8] Clinical risk prediction has been done with tool so, that patient takes care of his disease and he can undergo proper disease treatment steps. [9]. Clinical risk prediction helps clinicians to predict the chance of cardiac disfunctions (such as myocardial infarction, death). This proposed tool help in taking timely and appropriate intervention strategies to those at high risk of cardiac arrests and to motivate patients to remain alert following these strategies the patients can reduce high mortality. Extensive research work carried on clinical risk estimation of ACS. Earlier ACS risk-scoring tools like GRACE and TIM have been developed for monitoring population samples over long period. In proposed model patient's information is extracted from admission records of Hospital for clinical risk prediction models of ACS patients.

II. RELATED WORK

W are using SVM for the following reasons as in points [1][2][3][4[5]][6][7].[1].SVMs need less memory. SVM are less prone to errors. They perform text classification in HER systems and pattern recognition in echo-cardiograph imaging. [2].They help to stratify CV risk and physicians make decisions. SVM is feasible for large and complex data which is nonlinear, such as "omic" data. [3].SVM accuracy and enhance processing time. The wrong choice in kernel functions may lead to errors .SVM works even with small data sets. [4].The Decision Tree algorithm encounters over fitting because of the small datasets. It is used with a series of yes/no questions. SVM classify data into categories. [5].In addition it can be used in clinical decision-making. Random Forest algorithm is an extension to Decision Tree algorithm. In this algorithm Decision Trees are combined together
and every Decision Tree is trained independently.[6]. Random Forest algorithms compute angiography, readmission, HF risk and survival prediction models.

[7].The Naive Bayes classifier is probabilistic classifier obtained from Bayes theorem. It executes small training datasets which is be used in text classification problems in CV risk factor identification and decision-making systems. Fuzzy logic used in areas such as prediction of early coronary artery disease and mortality. K-nearest neighbor is nonparametric methods. It works on small training datasets. Fuzzy is used a in ECG interpretation problems. K-nearest neighbor takes more space and time when using large datasets.

III. EXISTING SYSTEM

There are many data mining algorithms like decision trees, fuzzy inference systems and Bayesian networks, etc., have been said to explore the uses of EHR data for clinical risk prediction. When considering an example Tay et al. gave an neural inspired deep learning ANN algorithm for risk prediction by using EHR data. There exists a Bayesian analyses which predict human clinical adverse events which very well helped in drug development programs. In existing work, a genetic fuzzy system designed for risk prediction in unstable angina. It also evaluated on a dataset which is collected from a Chinese hospital in EHR patients. We previously developed a probabilistic topic modeling. The said model finds risk stratification by exploring the potential of Electronic Health Recording unsupervised fashion.

IV. PROBLEM DEFINITION

Acute coronary syndrome (ACS), as a common and severe cardiovascular disease, is a leading cause of death and the principal cause of serious long-term disability globally. Deep learning architectures with a greater number of layers, can potentially extract abstract and invariant features for better performance of patient classification. The ability of inference on a large volume of heterogeneous EHR data is particularly suitable for our aim. Therefore, this paper proposes a novel approach for clinical risk prediction of ACS based on deep learning. Among various deep learning models, the Stacked Denoising Auto-encoder (SDAE) has particular advantages
such as rapid inference and the ability to reconstruct features. After considering all reconstruction learning phases, i appended a softmax regression layer on the top of the resulting reconstructed feature representation layer. This can help in solving clinical risk prediction .Our proposed model can learn more patient feature representations and improve the performance of clinical risk prediction.
The main objective of this research as stated below [1][2][3][4]. [1]Is to examine different deep learning algorithms in order to detect risk factors of coronary heart disease in a real-world, high dimensional Electronic Health Record data sets with heterogeneous data.[2]. The goal of the work is to find different algorithms and find algorithm that is suited for large data sets working.[3].In order for the chosen algorithms to scale different levels of risk factors in ACS with an increasing size of the patient sample size.[4].To find an algorithm that achieves Accuracy with less preliminary processing of the data.

V. IMPORTANT TERMS

This section briefly describes the background that is closely related to the proposed research paper. We discuss three issues here namely One is deep learning and its applications in biomedicine and the second one is clinical risk.

1. Deep learning

Deep learning is an Artificial Intelligence concept applying set of algorithms in machine learning that model very high-level abstractions in the data by using architectures which is
composed of multiple non-linear transformations.[1]. Stacking the nonlinear transformations is the basic idea used for deep learning algorithms.[2]. The more layers the data goes through within the deep learning architecture, the more tough to use nonlinear transformations construction. [3]. Stacked denoising auto-encoder (SDAE), is most investigated deep learning architectures. SDAE is a symmetrical neural network, which is used for learning the features from dataset in unsupervised fashion [4]. Auto-encoder in SDAE is trained to reconstruct input that is clean "repaired" from the version that is corrupted. [5]. Chen et al. proposed a deep learning Model for phenotyping from Electronic Health Records and the model is predictive modeling of chronic
diseases. [6], Tran et al. proposed a computational framework to harness EHR with a type of deep neural network. Many deep learning models promise results in learning good representations in an unsupervised manner. This model performs supervised tasks, such as clinical risk prediction etc. On the contrary, in proposed paper Deep learning

models using classes, it can reconstruct more discriminative features by using supervised task. Thus in the proposed model a regularized SDAE for clinical risk prediction of ACS has been developed in our research.

2. Clinical risk prediction

Clinical risk prediction models are used in healthcare for variety of applications as described in[1] to [3]..[1]. Mostly they are targeted to identify patients with higher risk of ACS disease. With levels like high, average and low. [2]. A model called GRACE risk scoring model, is a popular and well-accepted risk scores of ACS disease which was developed to predict clinical risk score on an individual patient each.

[3]. However, models along with this line have been estimated using a small set of especially of patient features from highly-stratified cohorts. [4] Bayesian networks and fuzzy inference systems have been proposed to find the potential of EHR data for risk prediction.[5] Karaolis et al., used C4.5 decision tree algorithm to retrieve essential risk factors of heart events from Electronic Health Records.[6] Hybrid model was developed to identify risk factors of heart disease in patients. My proposed model will adapt dynamic treatment information in EHR data to improve the performance of prediction for ACS model. It, and help in the clinical prediction of coronary heart diseases, from a large volume of EHR data.

3. CHALLENGES

Machine learning techniques using deep learning methods is new technology. They require properly organize data sets to provide correct answers to the questions we shoot them. A business depending on machine learning application needs to give time, resources, and take risks. Artificial neural network can take millions of parameters. A training set usually consists of thousands of records.ANN is capable of remembering the training set and it also gives answers with 100 percent accuracy.

Though it is giving answers to many questions it may prove useless when new data is given. This problem is known as over fitting (or overtraining). This limits deep learning algorithms usage.

VI. PROPOSED FRAMEWORK

The proposed framework contains modules like Administrator, Clinician and Researcher. The Administrator role is to make registration that is to add or remove users. The Clinician's module role is to find risk factors related to acute coronary syndrome symptoms. Disease regarded information is collected from samples of various patients and it is stored in MongoDB database. Researcher module role is to provide the parameters for the analysis and it has to find risk levels basing on predictions by applying Softmax regression algorithm to the provided data. Process flow of proposed framework can be seen verified in the below diagram which is given in proposed framework.

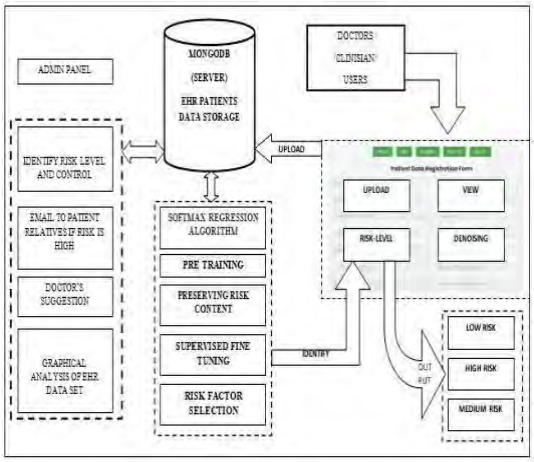


Figure No-1: System Architecture

Module Description

1. Pre training using SDAE

In this module we discuss the training and validation of ACS clinical risk prediction model by using our proposed approach. In SDAE model is a symmetrical neural network which is mainly used for learning all the features of a dataset by using techniques of unsupervised learning. So, as to build a deep learning architecture by using K hidden layers SDAE is trained using greedy layer-wise learning in unsupervised mode.

2. Preserving risk content into SDAE

SDAE has been used for classification problems the features reconstructed by SDAE are in unsupervised manner. While taking other side, risk information is contained in the training dataset. This dataset is included into learning so, as to make the reconstructed features that are beneficial for the future tasks of clinical risk prediction. We add two specific constraints like intra-risk-level affinity & inter-risk-level repulsion. These two constraints helps SDAE to preserve clinical risk information of training on patient samples. These two constraints

use the reconstructed feature representations of patients with the same risk level and also on different risk levels.

3. Supervised fine tuning

After the completion of pre training we added a Softmax regression layer. We placed it on the top of the reconstructed layer to construct a deep artificial neural network which is named as stacked denoising auto-encoder using Softmax regression model. RSDAE-S perform clinical risk prediction task. Here we fine tuned the RSDAE-SM using algorithm techniques of back propagation . This back propagation minimizes the cross entropy loss by taking help of Softmax regression layer.

4. Risk factor selection

In our proposed model risk factor selection strategy identify risk factors of ACS patients with different risk levels. We assume that patient features with less reconstruction errors are more re-constructible. Patient with more re-constructible features are likely to bear the below characteristics. When patient sample within a specific risk level there is small reconstruction error between the original and reconstructed data as it learn regularized SDAE model. We can also argue that different risk-level patient behave differently according to their risk factors, and produce a big reconstructed error. We expect a big reconstruction error in the discriminative learning. This happens as there is mismatch of input patient sample and the testing model used for implementation.

VII. ALGORITHM

Algorithum-1: Deep Learning Algorithms

Deep learning algorithms run its data using many "layers" of artificial neural network algorithms. Every layer passes its specified representation of the data to the other adjacent layers. All machine learning algorithms perform its tasks properly on datasets which have up to a few hundred columns or given features. However when taking an example of unstructured dataset,

like an image which include features that this process becomes completely unfeasible. An image with800-by-1000-pixel which supports RGB color has more than 2.4 million features. All these mentioned features are handled by several machine learning algorithms.

Algorithum-2:Support Vector Machine Algorithm

Support vector machine (SVM) is used to classify the patient in different categories. SVM is the algorithm that can assign two different problems like problem of classification and problem of regression. It is supervised learning algorithm. In this algorithm we have plotted each data item as a separate point in n dimension plane. Where n is the number of features, in our case we have taken twelve major features who are considered as cause of ACS.SVM areconsidering co-ordinates of individual observation from the report of patient. In our system we have considered three hyper planes to classify patient in three different categories. we have kept maximum possible distance in each hyperplane so that the category of each patient should get accurately classified. If we select less distance between two hyper planes then there is high chance of miss classification between the patient. Many methods are available for classification of data but our analysis shown that SVM is giving best accuracy i.e. more than 90% accuracy in classification.

VIII. SOFTWARE DETAILS

Used software and algorithms in this research are [1][2][3][4][5][6].[1].Operating System used in research is Windows -7. [2]. Application Server used in research is Tomcat Server 8.0 [3]. Front End details are JSP, Java. [4].Database that is used is MongoDB. Database Connectivity used is Robomongo-0.8.5-i386. MongoDB is used because, MongoDB uses its storage pattern in the

MongoDB provides redundancy and high availability. Data can be distributed across multiple servers. MongoDB is used for this research work as it is supporting NOSQL. It can deal with BIGDATA analytics. As the Electronic Health records are huge volumes of data and related with machine learning technique.[5] Javais used for front end design as the data is more secure and robust. Since GUI design is possible in java it is opted for my research work.

IX. EXPERIMENTAL DATA

Several experiments that were conducted to compare the performance of the proposed model by using "Start-of-the-art" methods on the clinical risk prediction task.

A. DIAGONOSIS OF ACS

The study sample consisted of more than 2000 patients admitted to the General Hospital with a diagnosis of ACS. ACS was said as symptoms of Ischemia along with positive cardiac enzymes, new

coronary Artery disease or previous existence of Cardiac Artery Disease. In an early stage of hospitalizations of ACS patients, we gathered the 1experimental dataset from admission patient's record sets of the EHR system as shown in figure -2

Figure No-2: Different field of EHR Data set Preprocessing was done on the datasets. Patient features mapped with binary values into 0 and 1. Feature "Gender" can be mapped as 0 assigned to "male"& 1 Assigned to "female". Patient's features with multiple categories were split into multiple categories of variables. Cardiac function is given 1, 11 and 111 level rating basing on severity. Patients Features were stacked into datasets with minimum and maximum value. Top 20 patients feature of ACS were recorded. Basing on high, low, medium patients risk levels of ACS records were stratified.

B. AUC

Accuracy is used as a method for the prediction performance. Medical data-sets when used the accuracy can be high, when classifying the Minority samples with that of majority sample class. Patients at low-risk levels are much greater when compared with that of high that of high-risk levels. Receiver operator characteristic curve (AUC) is used. The receiver operatorcharacteristic curve plots true positive rate against the false positive rate (specificity). A good method produces Higher AUC values for ACS risk prediction. A typical measure of prediction accuracy was plotted to find high-risk cases varied with the number of low risk patient cases in the sample.

C. Experimental Study

We have taken the generalized data set of 2000 patient record of all different fields. Every field of HER data set is very important for processing the records. Approval and suggestion was taken from cancer specialist doctors and institutional review department to carry this research work.

D. Comparison with state-of-the-art supervised learning algorithms

We compared the performance of the proposed RSDAE-SM model with the regular SDAE adding a softmax layer (SDAE-SM). We also

used five state-of-the-art classification algorithms, i.e., Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Multi-layer perceptron neural networks (MLP), and naive Bayes (NB), as benchmark methods.

E. Informative risk factor selection

The proposed model is used to find informative risk factors from Electronic Health Data Records data. Top 10 risk factors for the patient group with same risk. Level were taken and confirm their validity with the clinical experts, patient features were divided into two classes 0 or 1 (saying patient case has this particular mentioned feature or not) in the model learning process. Creatinine, smoking status, etc validated inside the clinical cohort study. To apply this proposed model in clinical practice, we should consider the

simpler model implementation which investigates the simplicity and accuracy. Clinicians are interested to know risk factors that collect to obtain a predictive performance. We apply our risk factor selection strategy to find risk factors with regard to each risk level. We use top 20 in the union of these risk factors to proposed model for our ACS risk prediction model.

X. RESULT ANALYSIS

We have analysis this work on the basis of low, high, medium risk category. How many numberof patient in each category which shown in graphical format.

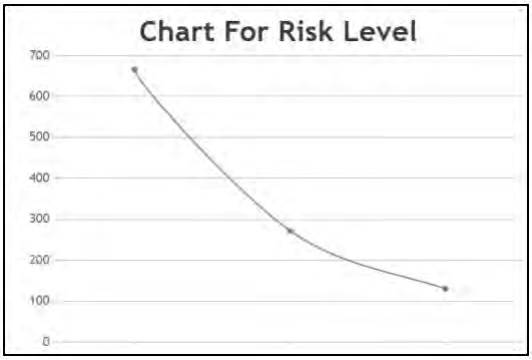


Figure No-3: Spline Chart Analysis

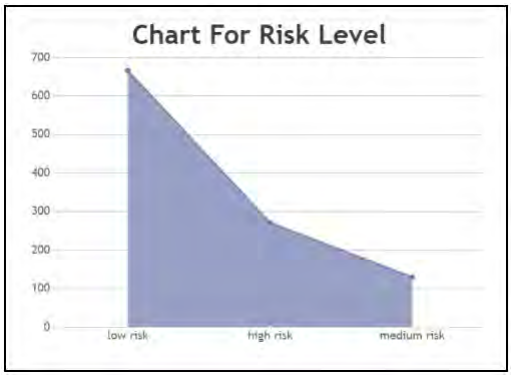


Figure No-4: Area Chart Analysis

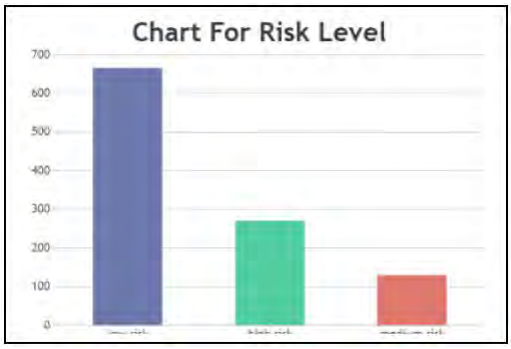


Figure No-5: Bar Chart Analysis

Above chart shows number of risk patients categorized in different risk levels. Depending upon the input values of patients different test records, system assigned various risk levels. Maximum patients who gets admitted as uneasy or restless feeling are coming in low risk category. Some patients who are not having any complaints initially, but their various test shown that they are at high risk and need to be treated quicky.

XI. CONCLUSIONS

This work proposes learning approach to discuss the clinical risk identification problem of ACS from Electronic Health Record data. [1].In Existing ACS risk methodologies relied on asmall set of known risk factors, in this proposed approach we are able to utilize a large volume of heterogeneous Electronic Health Records.[2]. Our Model is an accurate and robust clinical risk prediction model which gives the 90% accuracy to predict the level of cancer patients using the different data set. In this work we have used SVM algorithm which achieved the best result also our model directly informed to patient relatives for caring of patients if the risk is too high and they got good suggestion from expert doctors. The experiments in the research were conducted on a clinical dataset and the results shows that the proposed model is able to achieve good performance in clinical risk prediction, when compared with the algorithms like state-of-the-art classification algorithms.

References

[1] D.C. Goff, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report
of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation, 2014, 129, S49-S73.

[2] E. Boersma, et al. Predictors of outcome in patients with acute coronary syndromes without persistent ST-segment elevation. Resultsz from an international trial of 9461 patients. Circulation, 2000, 101(22):2557-2567.

[3] P.W. Wilson, et al. Prediction of coronary heart disease using risk factor categories. Circulation, 97:1837-1847, 1998.

[4] Z. Huang, et al. A probabilistic topic model for clinical risk stratification from electronic health records. Journal of Biomedical Informatics, 2015,58:28-36.

[5] M. Matheny, et al. Systematic review of cardiovascular disease risk assessment tools. Technical Report, Agency for Healthcare Research and Quality (US), 2011.

[6] Minas A. Karaolis, et al. Assessment of the Risk Factor of Coronary Heart Events Based on Data Mining With Decision Trees, IEEE Transactions on Information Technology in Biomedicine, 2010, 14(3):559-567.

[7] S. Bandyopadhyay, et al. Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data, Data Mining and Knowledge Discovery, 2015, 29(4):1033-1069.

[8] S. Paredes, et al. The CardioRisk project: Improvement of cardio vascular rrisk assessment. Journal of Computational Science, 2015,9:39-44. [9]Jessica L. Mega, et al. Rivaroxaban in Patients with a Recent Acute Coronary Syndrome. The
New England Journal of Medicine, 2012,366(1):9-19.

[10]Elliott M. Antman, et al. The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. The Journal of the American Medical Association, 2000, 284(7):835-842.

[11] Shaun G. Goodman, et al. The expanded global registry of acute coronary events: baseline characteristics, management practices, and hospital outcomes of patients with acute coronary syndromes. American Heart Journal, 2009, 158(2):193-201.

[12] C.J.L. Murray, A.D. Lope. Global mortality, disability, and the contribution of risk factors: global burden of disease study, Lancet, 1997,349 (9063):1436-144

Study on Analysis of Gene Expression Dataset And Identification of Differentially Expressed Genes

Saroj Ghadle, Rakesh Tripathi, Sanjay Kumar, Vandana Munde
N. I. T. Raipur, Chhatisgarh, TCET College,
Mumbai (M.S.)
saroj.shambharkar@gmail.com

Abstract. **The diagnosis of complex genetic diseases such as cancer, cardiovas-cular, respiratory diseases, human disease resistant genes(R-Genes) in Homo sapiens at early stage is a need is a challenging task in the today's biomedical and other research field. In this paper the analysis of gene expression dataset of is done and discussed. The gene expression dataset is considered because of it high dimensional data and gives more scope for analysis of diseases. The gene expression dataset E-GEOD-6575 using heatmap and tumor gene expression dataset using classification technique is analyzed. The classification of gene ex-pression dataset of 801 samples of RNA-seq tumor (cancer) is discussed. Each sample is associated with set of gene expression values. And based on these gene expression values the samples are distributed among the 5 different types of classes of tumors. The main objective of paper is to understand the gene ex-pression dataset of diseases and purpose of analyzing it. The change in gene ex-pression values considered as differentially Expressed Genes. The identification of differentially Expressed Genes will be helpful for good and poor prognosis of disease.**

Keywords: Classification, Differentially Expressed Genes, GEO dataset, gene expression, Heat map, National Centre for Biotechnology Information, NCBI.

I. INTRODUCTION

There are two technologies, microarray and RNA-seq technology used to generate the gene expression data. The paper objective is to understand the importance of anal-ysis of gene expression dataset. There are various techniques and methods used to visualize, interpret and understand the gene expression data. The objective behind the visualization, interpretation and understanding gene expression data is to get the knowledge about the condition of a human being. The analysis is helpful for the phy-sician for proper diagnosis of the disease.

The National Centre for Biotechnology Information, NCBI website provides us the gene expression dataset .The required gene expression datasets can be searched through Gene Expression Omnibus (GEO).From NCBI, the Gene Expression Omni-bus, GEO dataset E-GEOD-6575, gene expression dataset is a imported from NCBI. The domain experts have uploaded their data after conducting experiments on various diseases cases, and they are freely downloaded by the user. These dataset are authentic and regularly updated.

1.1 Gene Expression dataset

The gene expression tumor dataset downloaded from UCI Machine Learning Re-pository. This dataset consisting of 801 samples of 5 different types of tumor with 20531 attributes. These attributes associated with gene expression values. Before the analysis of gene expression dataset the dataset has to go through the preprocessing depending of dataset considered by us. The gene expression dataset sometimes may get affected by the certain environment conditions. And because of change in envi-ronment conditions the noise, missing values and inconsistency found in the dataset, such dataset requires the preprocessing.

As we mentioned, the analysis is performed on the datasets of tumor and E-GEOD-6575 of gene expression in blood of children with autism. The dataset of gene expres-sions fetch from the website National Centre for Biotechnology Information, NCBI and UCI is used. The gene expression dataset of cancer (tumor) and gene expression dataset with accession number E-GEOD-6575 is considered and studied. The paper focused on analysis of change in gene expression attribute values associated with each sample.

II. LITERATURE SURVEY

Earlier in past decades,linkage analysis was carried out for the purpose of identification of novel disease genes.The problem or disadvantage of linkage analysis is it is time consuming and expensive method.In the research papers it is mentioned the diagnosis of Autism Spectrum Disease is costly and time Consuming Process using linkage analysis,so to overcome this problems the machine learning approach can be used.In the literature the authors used a Random Forest Classifier ,and they achieved

80% accuracy for predicting new Autism Spectrum Disorder(ASD) genes and their results are better than previous work.In this paper for identification of biological markers the machine learning methods were used,and they used Gene Ontology to improve the identification of biological markers in complex diseases like ASD.The training performed by machine learning classifier done using Gene Ontology.And performance can further be improved by incoorporating functional similarities.There are algorithms like Random Walk Restart(RWR) which are widely used for prediction of disease genes.The gene expression helps the physician to know the disease transition in human body due to change in Gene expression profile.

In paper "Prediction of gene function by genome-Scale Expression Analysis: Prostate cancer-Associated genes" the authors objective was identifying the diseases associated with genes and discovered for cancer using the Guilt By Association method. In this paper they worked on the expression pattern consist of 40,000 human genes taken from 522 cDNA libraries.From the survey papers it is understood that the disease similarity network can be extracted from a disease similarity matrix,using text mining techniques performed by taking OMIM records**.**But the accuracy of this approach is less and it is improved if the we use phenotype ontology Database and also improves prediction performance.

The network of protein–protein interactions also help and provides us the valuable insights into the inner workings of cells. Accordingly, it is crucially important to develop an automated method or high-throughput tool that can efficiently predict the PPIs. The authors of paper has introduced a predictor, called iPPI-PseAAC(CGR), was developed by incorporating the information of "chaos game representation" into the PseAAC (Pseudo Amino Acid Composition). The predictor has used the machine learning algorithm ,Random forest.

The advantage of this predictor mentioned by them is the key sequence-order or sequence-pattern information obtained can be more effectively incorporated during the treatment of the protein pair samples. It has been observed by them using the cross-validations applied on their datasets given the remarkably higher success rate than the existing.

The following challenges identified after going through literature survey and study:

(i) In post-genomic scenario, the challenge is to identify and predict "disease genes" from the large amount of genetic related information

(ii) Identify biological marker from complex diseases is difficult as these diseases presents highly heterogeneous genotype

(iii) Identify and predict human disease Resistant-genes from large genomic data

(iv) Identify and predict disease and non-disease genes from genomic dataset or

predict true disease genes and essential genes from genetic dataset.To solve the above mentioned challenges there is a need of some techniques,computational methods,approaches or tools.The question comes from the study for complex diseases how we can analyze and identify the gene expressoins are changed or not.And what is the purpose of analyzing gene expression datasets.From the literature it is observed that the researchers employed varoius techniques to study and analyze the genes related information,some of them used machine learning approach and deep learning approaches and concluded they are effective, powerful computational technique to understand and analyze the genomic datasets.

2.1 Visualization and Interpretation of Gene Expression dataset

The gene expression dataset E-GEOD-6575 is analyzed using heat map which is combined with clustering. The most common method to visualize the gene expression dataset is heatmap. The heat map is a grid like structure containing rows and columns, the rows indicating genes and columns indicating samples. The color and intensity of each grid indicates the state of gene expression. From the color and intensity we can analyze that there is change in gene expression.

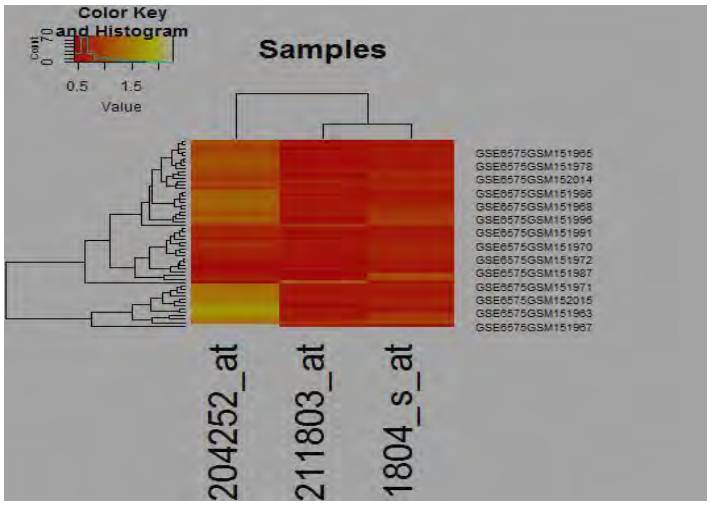


Fig.1. A figure showing the visualization of gene expression dataset GEOD-6575 using heat map

The heat map of E-GEOD-6575 gene expression dataset is obtained using R-programming shown in the figure. The heatmap shown is containing the group of gene expressions patterns based on the similarity. Therefore the most similar gene expression patterns were grouped together.

The figure 1 showing values using different colors in each sample. The colors yel-low and red used to analyze the attribute values of E-GEOD-6575 gene expression. In the fig 1 there are 3 samples (represented using 3 columns) containing probe sets 204252_at, 211803_at, 1804_s_at as attributes or considered as samples. The rows of heatmap consist of gene expression values of genomic series GSE6575 dataset. In middle column (sample specified by probe set 211803_at) we can observe the similar-ity is more and they used same color. There is less change in gene expression values more are in red color and less values in yellow colors. It means less change in gene expression values in middle sample. But if we observe the sample specified by probe set 204252_at and 1804_s_at, there is more variation in values for same set of values (values specified by rows) and less variation in sample 1804_s_at as compared to 204252_at.If the change in expression values is more we conclude them as differen-tially expressed. So, among 3 samples we conclude more changes or differentially expressed genes in first leftmost sample (204252_at).

In this way heatmap are giving good visualization and analysis of values stored E-GEOD-6575 gene expression dataset.

2.2 Classification of RNA Seq. Gene Expression Dataset of cancer

The Gene expression dataset for cancer RNA-Seq dataset is fetched from the UCI repository (Machine Learning).The the size of the dataset is 201.354 MB. The 801 samples were taken and each sample containing the change in gene expression value. Each sample consists of numerical values of gene expressions. There are total 20531 attributes from gene_0 to gene_20530. This dataset also giving us different way to analyze the classification of each sample based on values of associated with each sample.

The figure 2 is showing the classification of 801 patient samples RNA-Seq PANCAN dataset. The dataset referred by us is used by the authors Weintenstein, John N. et al. in their paper 'The cancer genome atlas pan-cancer project 'published in Nature genetics 45.10 (2013).

The patients are classified into 5 classes of different types of tumors. The patient is classified into different categories of tumors and found more cases under category of BRCA in the dataset of 801 samples of tumor patients. The classification tells us the 801 samples containing 300 patients suffered from BRCA, 146 with KIRC, 141 of LAUD, 136 with PRAD and only 78 with COAD tumor. This helps a physician to analyze the patients with BRCA tumor, what makes changes in gene expression and how it can be reduced. From the samples of 801 patients we conclude from the classi-fication more changes in gene expression values in samples of BRCA and conclude more differentially expressed genes in it.

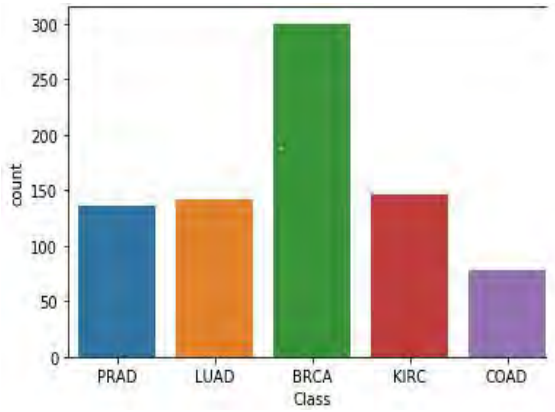


Fig.2. A figure showing the classification of RNA-seq gene expression dataset of cancer.

This classification done for the patients whose changed expression values changed and the change in gene expression shows the patients are suffered by cancer. From the classification results we say the gene expression dataset analysis gives us proper deci-sion on the available sample of a patient.

III. CONCLUSION AND FUTURE WORK

In this paper the main focus is on understanding the gene expression dataset and change in values of gene expression. And change in gene expression makes them differentially expressed. It is studied from the literature too that the identification of differentially expressed genes will help in analyzing the state of a patient.

The analysis of gene expression dataset, we can identify the differentially ex-pressed genes. From the literature also it has been observed that the identification of the differentially expressed genes requires the comparison of gene expression levels or read counts at different conditions of an organism say human. The difference be-tween gene expression level at normal condition and abnormal condition has to be calculated to diagnose the disease. If the difference is found then we conclude the gene expression levels have been changed and the patient condition is not normal. In future work we will work out the analysis of gene expression dataset using different Machine Learning algorithms to identify differentially expressed genes. The analysis can also be done using python programming language on gene expression dataset of diseases to identify differentially expressed genes.

References

1. Baron, C.A., Liu, S.Y., Hicks, C., Gregg, J.P.: Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. J Autism Dev Disord 36, 973–982(2006).

2. Bram van den Bekerom, 2017. Using Machine Learning for detection of Autism Spectrum Disorder.*26th twente Student Conference on IT Feb 3th 2017*, Enschede ,Netherlands 2(5), 99–110 (2016).

3. Chanda Panse , Manali Kshirsagar, Dhananjay Raje , Dipak Wajgi: Methods used for

Identification of Differentially Expressing Genes (DEG's) from Microarray gene Dataset: A Review. International Journal on Bioinformatics & Biosciences (IJBB) Vol.6, No.1, March 2016 DOI: 10.5121/ijbb.2016.6101 (2016).

4. Ching Siang Tan, Wai Soon Ting,Mohd Saber iMohamad,Weng Howe Chan, Safaai Deris,

and Zuraini Ali Shah, A Review of Feature Extraction Software for Microarray Gene Ex-pression Data. Hindawi Publishing Corporation BioMed Research International Volume 2014, Article ID 213656, 15 pages http://dx.doi.org/10.1155/2014/213656.

5. Jianhua Jia , Xiaoyan Li, Wangren Qiu , Xuan Xiao , Kuo-Chen Chou:.iPPI-PseAAC(CGR): Identify protein-protein interactions by incorporating chaos game repre-sentation into PseAAC. J Theor Biol. 2019 Jan 7;460: 195-203. doi: 10.1016/j.jtbi.2018.10.021. Epub 2018 Oct 9(2019).

6. Eunji Jun, Eunsong Kang, Jaehun Choi, Heung-Il Suk: Modeling regional dynamics in low-frequency fluctuation and its application to Autism spectrum disorder diagnosis, *Neu-roImage 184*, 669-686 (2019).

7. E Skafidas, R Testa, D Zantomio, G Chana, IP Everall, C Pantelis: Predicting the diagno-sis of autism disorder using gene pathway analysis, 2014.*Molecular Psychiatry*. 19,504-510(2014).

8. Gourab Kundu , Sirajum Munir, Md. Faizul Bari, Md. Monirul Islam, and K. Murase: A
Novel Algorithm for Associative Classification",1*4thInternational Conference, ICONIP 2007*, Kitakyushu, Japan, pp 453-459 , November 13-16 (2007).

9. Jason S. Nomi, Lucina Q. Uddin: Development Changes in large-scale network connec-tivity in autism, *NeuroImage Clinical 7,*732-741(2015).

10. Joanna S. Amberger, Carol A. Bocchini, Franc¸ois Schiettecatte, Alan F. Scott and Ada Hamosh: OMIM.org: Online Mendelian Inheritance in Man (OMIM @), an online cata-log of human genes and genetic disorders. *Nucleic Acids Research, 2015, Vol. 43, Data-base issue D789–D798 doi: 10.1093/nar/gku1205* (2015).

11. F., Janet Pinero, Alex Bravo, Nuria Queralt-Rosinach, Alba Gutierrez-Sacristan Jordi Deu-Pons, Emilio Centeno, Javier Garcia-Garcia, Ferran Sanz and Laura I Furlong: Dis-GeNET: a comprehensive platform integrating information on human disease-associated genes and variants Nucleic Acids Research, 2017, Vol. 45, Database issue D833–D839 doi: 10.1093/nar/gkw943(2017).

12. Megu Ohtaki,Keiko Otani, Keichi Satoh,Toshihiko Kawamura,Keiko Hiyama and Masahiko Nishiyama: Model=based analysis of microarray data: Exploration of Differen-tially expressed genes between two cell types on a two-dimensional mixed normal model: Japenese Journal of Biometrics Vol.26, No.1,31-48 (2005).

13. Andrea Bisognin, Stefania Bortoluzzi and Gian Antonio Danieli : Detection of chromoso-mal regions showing differential gene expression in human sketal muscle and in alveolar rhabdomyosarcoma:BMC Bioinformatics,pp. 5-68 (2004).

Review Paper on E-Commerce and Recommendation System

Dr. Nita Thakre,Vedant Santoshwar,Aditya Upadhye,Prashant Roy,Aakash Shende, Pratik Paraye
Department of Computer Technology
Priyadarshini College of Engineering, Nagpur
bfbiznet@gmail.com, bfbiznet@gmail.com, bfbiznet@gmail.com

Abstarct **- In today's world, AI is getting better and better in predicting people's behaviour more efficiently and accurately. E-commerce is no exception. Right from finding the shortest path possible to deliver the product, to providing more relatable suggestions to the user, AI is integral to e-commerce business model. Recommendation system of the e- commerce websites/applications is the best example of AI at work. Recommendation system processes the user data and provides the results that the user is searching for. It has made the search engine of the e-commerce websites/applications more powerful and to the point, filtering out the results that the user may not be interested in.**

Keywords - BFBiz.net,E-commerce, website, Online shopping within the city, Delivery, Electronic

I. INTRODUCTION

E-commerce is all about selling and buying of goods and services over the internet. Before e-commerce, it was done without internet physically in the markets but after the arrival of e-commerce in India our life has become more comfortable because of its number of advantages. The services offered by e-commerce are online shopping of anything 24x7 and at any place, customers can find the products on e-commerce websites which is not available in physical markets, it reduces cost and time, without stepping out of home we can get our product at home. But, due to e-commerce the revenue generated by local vendors is decreasing day by day. People prefer to the offers provided by the online stores to save money by comparing different stores. The latest report says that India has approximately $16.8 billion Mobile Commerce market size and mobile commerce is set to become the primary way to shop online in India. It is already used for 46% of transactions, one of the highest rates in our report series. Mobile commerce will expand at a compound annual growth rate of 31.2%, to reach a value of $49.8 billion by 2021. To deal with the problems of local stores we are making an E-commerce website “BFBiz.net” where the required products will be shown from the nearest store within the city from the user’s location by comparing the prices of different stores within the selected distance. This will make the goods sell and the delivery can be done very fast within the same day.

II. LITERATURE SURVEY

The concept of recommendation systems emerged in mid- 90s. In past 15 years there has been a tremendous boom in the development of recommendation sites. The people using the recommendation systems is increasing exponentially making it crucial for these systems to generate recommendations that are close to the items of user’s interest. Micheal Pazzani [1] proposes recommending data sources for news articles or web sites after learning the taste of the user by learning his profile data. This paper mentions information that can be considered to learn the profile of a user. Based on ratings given by a user to different sites, ratings that other users have given to those sites and demographic information about users the recommendations can be made. This paper describes how all the processed information
can be combined to provide recommendations to the users.

Balabanovic, M. and Shoham[2] discuss that 'personalized recommendation systems' are widely used in e-commerce websites to provide users with recommendations. It suggests that Hybrid approach is the best possible way for recommendation to user rather than using only content-based filtering or only collaborative based filtering.

Satya Prakash Sahu, Anand Nautiyal, Mahendra Prasad [3] presented a best possible precision and have presented a comprehensive comparative analysis by selecting Content Based Filtering, Collaborative Based Filtering, Hybrid Content-Collaborative Based Filtering, k- mean clustering and Naive Bayes classifier among various machine learning techniques which can be used to conclude the recommendation system.

III. PROS & CONS

- Highly relevant: As the system uses the preferences of the user, it is able to show only the relevant data which the user may may find interesting. It is very helpful in narrowing down the search results for that user.

- Transparency: Transparency is another important aspect of Content- Based filtering. As the user can know the reasons for which the products are recommended, it makes the system more trustworthy. As oppose to the Collaborative-Based filtering in which products are recommended by preference of other users. The process of recommending the products is open to the user.

- User Friendly: While the system waits for some initial inputs, it is relatively faster in recommending the items. Gradually over the time, after hundreds or thousands of search results it recommends the products more precisely and accurately to the user.

- New items can be recommended more quickly: In Collaborative-Based filtering new items may not be recommended immediately as the users may not have look for that item. As oppose to this, in Content-Based filtering items can be recommended more quickly. Content-Based filtering does not require the users to interact with the object in order to recommend them as it is based on individual preferences.

IV. PROPOSED SYSTEMPHP, SQL, Bootstrap, HTML, CSS, JavaScript, jQuery, Ajax, Machine Learning

–Proposed Methodology

In this the user will be shown the products from its nearest locations within the same city. The products will be recommended according to the search history of the user. The products will be given discounts by the seller on his personal choice based on profits of both seller and customer. Here are languages that we have used and underlying reasons:

- PHP – Open Source, ML libraries easily available also it is probably the most widely adopted programming language. It's a highly scalable, functional, and object oriented language.

- SQL – Communication with Database, SQL is a lightweight, declarative language that does a lot of heavy lifting for the relational database, acting like a database's version of a server-side script. One particular advantage of SQL is its simple-yet-powerful JOIN clause, which allows developers to retrieve related data stored across multiple tables with a single command.

- Bootstrap – Lightweight and customizable. Responsive structures and styles hence we have mobile ready website

- HTML – For document design and display on web page

- CSS – for Designing Web pages and make them user friendly

- JavaScript/jQuery – for Client-side interaction that works on almost every web browser. It also makes websites launch super-fast and provides the end users with enhanced user experience (UX).

- Machine Learning – It is used for recommending products to the users based on their search history and one's interest.

- Ajax- It is used to create dynamic webpages where only part of whole web page is updated, if needed, without reloading entire webpage.

V. CONCLUSION

The other recommendation systems use historic data of the current user and user data of other users with same preferences, while our current system requires less historic data and operates on real time using the inputs of current user. The delivery system is also faster, as it delivers the package within few hours.

VI. FUTURE PLAN

To implement the system in different cities of India for the facilitation of the local shops and to enhance the delivery system using Honeycomb Structure where each cell will represent an area and will have its own delivery person to hand over the parcel to another cell to reach the final destination.

References

1 "A Framework for Collaborative, Content-Based and Demographic Filtering". Micheal Pazzani, University of California, 2004.

2 Balabanovic, M. and Shoham, Y. (1997).Fab:Content-Based, Collaborative Recommendation. *Communications of the ACM* 40(3), March 97.

Deep Learning Techniques for Detecting Deepfake: A Review

Abstract: **Deepfake algorithms can make forged pictures and videos that people can't differentiate them from true ones. The suggestion of technology that locate and prove the truth of virtual visual media is as a result essential. This paper offers a survey of tools and algorithms used to make deepfakes and, additional significantly, methods to locate deepfakes. We present large discussions on challenges, studies developments and guidelines related to deepfake technologies. By reviewing the history of deepfakes and cutting-edge deepfake detection strategies, this gives a comprehensive assessment of deepfake techniques and helps the development of latest and more robust strategies to deal with an increasing number of tough deepfakes.**

Keywords: deepfakes, deep learning, autoencoders, genereative adversarial networks (GAN).

I. INTRODUCTION

Deepfake (stemming from "deep learning" and "fake") is a technology which creates fake images or videos of targeted humans by swapping their faces another character saying or doing things that aren't absolutely done by them and humans start believing in such fake, as it isn't always recognizable with the everyday human eye. Deep learning models such as autoencoders and generative adversarial networks have been applied widely in the computer graphics vision to solve various issues [1,2]. These models have been used by deepfake algorithms to examine facial languages and actions of a person and synthesize facial images of another person making equivalent expressions and movements [3]. Deepfake algorithms usually require a large dataset to train models to make realistic images and videos. As public figures such as legislator or celebrities may have a large number of videos and images available online, they are probable targets of deepfake. Deepfakes were used to superimpose faces of legislator or celebrities to bodies in porn images and videos.

The first deepfake video found in 2017 where face of a celebrity was superimposed to that of a porn actor. It is threat to world security when deepfake procedures can be employed to make videos of world leaders with forged speeches for false purposes. Deepfakes therefore can be abused to cause political or religious misunderstanding between countries, to fool public and affect results in election campaigns, or create confusion in financial

markets by creating fake news. It can be even used to

create forged satellite broadcasting images of the Earth to hold items that do not really exist to create chaos in military**.**

There is also advantages of deepfakes such as creating voices of those who have lost theirs voice or updating episodes of web series without reshooting them . However, the number of venomous uses of deepfakes largely dominates that of the positive ones. The growth of advanced deep learning networks and the accessibility of big amount of data have completed the forged images and videos almost unique to people and even to algorithms. The method of creating those forged images and videos is also much easy today as it wants as little as a self-photo or a small video of a targeted individual. Less effort is required to produce an impressively substantial tempered footage. Recent advances can even create a deepfake with just a motionless image. Deepfakes therefore can be a threat affecting not only community figures but also ordinary people.

Finding the fact in digital domain therefore has become gradually critical. It is even more interesting when dealing with deepfakes as they are majorly used to help venomous drives and almost anyone can create deepfakes these days used with present deepfake tools.

Therefore, here have been abundant methods proposed to detect deepfakes. Most of them are also based on deep learning, and thus a battle between malicious and optimistic uses of deep learning methods has been rising.

In Section 2, we present the ideologies of deepfake algorithms and how deep learning has been used to enable such riotous technologies. Section 3 reviews different methods for detecting deepfakes as well as their advantages and disadvantages. We discuss challenges, investigation movements and directions on deepfake recognition as well as interactive program and forensics issues along with research gap in Section 4. Section 5 concludes the paper.

II. DEEPFAKE CREATION

Deepfakes have become popular due to the quality of tampered videos and the easy-to-use skill of their applications to a wide range of users with various computer skills from expert to beginner. These applications are typically developed based on deep learning practices. Deep learning is well known for its ability of instead of complicated and high dimensional data. The first attempt of deepfake creation was Fake App, developed by a Reddit user using autoencoder-decoder blending structure. In that method, the autoencoder extracts hidden features of face images and the decoder is used to reconstruct the face images. To swap faces between source images and target images, there is a need of two encoder decoder sets where each pair is used to

train on an image set, and the encoder's parameters are joint sandwiched between two network sets.

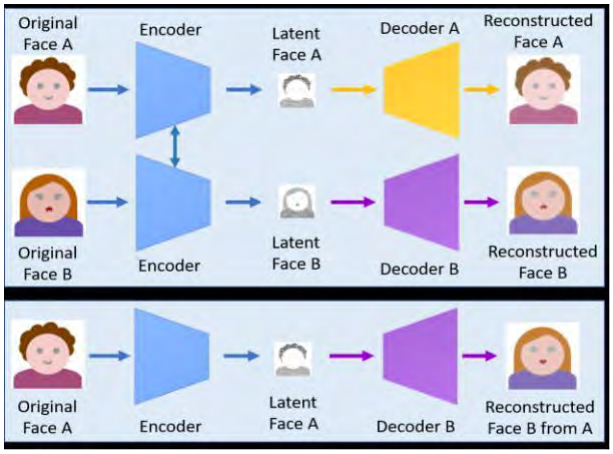


Fig. 1: A deepfake creation model.

This approach enables the common encoder to find and learn the comparison between two sets of face images, which are comparatively unchallenging because faces normally have similar structures such as eyes, nose, mouth places. Fig. 1 demonstrations a deepfake making procedure where the feature set of face A is linked with the decoder B to rebuild face B from the original face A. Table 1 shows some tools for deepfake creation along with its algorithm used and key features

Table 1: Summary of notable deepfake tools

Tools	Algorithm used	Key features,
DFaker	GAN with DSSIM	Reconstruction of face is done by DSSIM loss function. Keras library-based implementation
Faceswap-GAN	GAN v2.2 with MTCNN	the auto-encoder architecture is provided with adversarial loss and perceptual loss (VGGface) are added to.
Faceswap	GAN with MTCNN and DFL-H128	Parameters of the encoder are shared and pairs of encoder-decoder.
DeepFaceLab	GAN-er with MTS3FD	Expand from the Faceswap model with new models, Uses multiple face
DeepFake-tf	GAN with DSSIM	Like DFaker but tensor-flow is used for implementation.

This segment gives a review of deepfake detection methods where we categorize them into two major classes: fake image detection approaches and fake video detection approaches as shown in Fig. 2. The latter is distinguished further: visual artifacts within video frame-based methods and temporal features across frames- based ones. Whereas most of the approaches based on temporal features use deep learning recurrent classification models, the methods use visual artifacts within video frame can be executed by either deep or shallow classifiers.

III. DEEPFAKE DETECTION

Deepfake finding is normally deemed a binary arrangement problem where classifiers are used between reliable videos and interfered ones. This kind of method requires a large database of real and fake videos to train classification models. The number of fake videos is progressively available, but it is still limited in terms of setting a level for validating many discovery methods.

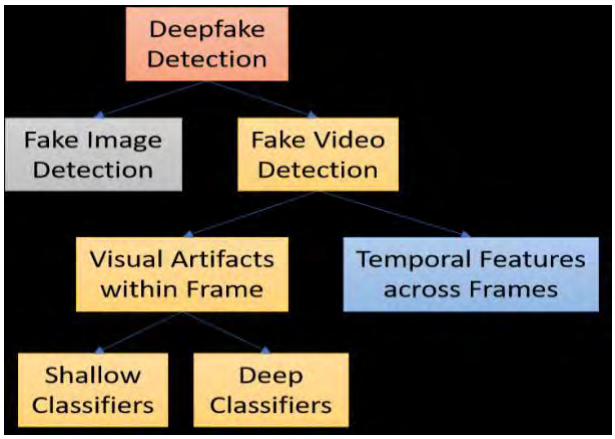


Fig. 2: Classification of deepfake detection.

.

To report this matter, Korshunov and Marcel [6] produced a distinguished deepfake data set containing of 598 videos based on the GAN with the open source code Face swap-GAN. Videos from the openly available VidTIMIT database were used to produce low and high quality deepfake videos, which can efficiently emit the facial features, mouth actions, and eye blinking. These are then used to test various deepfakes. Test outcomes show that the general face recognition systems supported VGG and Facenet [6] are incapable to detect deepfakes successfully. Additional approaches such as lip-syncing and image quality metrics with SVM makes very high error rate when allied to detect deepfake videos from this newly shaped dataset. This increases fears about the serious need of upcoming development of more strong systems that can detect deepfakes from the original.

3.1 Fake Image Detection

Face swap has a few convincing applications in video compositing, transformation in portraits, and specially in individuality safety as it can swap faces in photos by ones from a group of typical images. However, it is one of the approaches that fake attackers enter authentication structures to gain access. The usage of deep learning such as CNN, GAN, SVM, Random forest and multi-layer perceptron has exchanged face images more challenging for forensics as it can reserve pose, facial appearance and light of the photos. Among deep learning produced images, those produced by GAN are possibly most tough to notice as they are genuine and good quality based on GAN's ability to learn supply of the input data and gives novel outcomes with same input distribution.

Maximum works on recognition of GAN produced images do not consider the capability of the recognition models although the development of GAN is continuing, and
several new extensions of GAN are often introduced. Xuan et al. [7] used an image preprocessing phase. This rises the pixel level statistical comparison between actual image and false image and needs the forensic classifier to study additional essential and meaningful features, which has better simplification competence than preceding image forensics methods or image stag analysis networks.
Additionally, Agarwal and Varshney [8] company the GAN-based deepfake recognition as a hypothesis challenging issue where a statistical outline was presented by means of the information theoretical study of verification. The least distance among deliveries of genuine images and images produced by a specific GAN is defined, explicitly the vision fault. The logical outcomes demonstrate that this distance rises when the GAN is fewer correct, and in this situation, it is easier to distinguish deepfakes. In case of high-resolution image, a very accurate GAN is required to produce fake images that remain tough to detect.

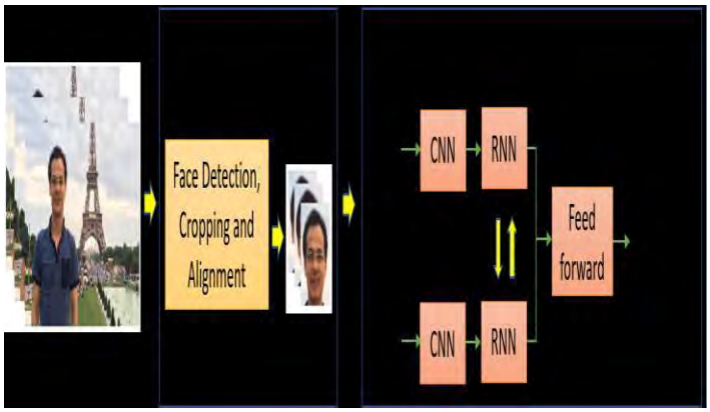


Fig. 3: A two-step process for face manipulation detection.

3.2 Fake Video Detection

Maximum image based deepfake recognition approaches cannot be used for videos because of the robust deprivation of the frame data after audiovisual compression. Also, videos have chronological features that are diverse amongst sets of frames and thus challenging for approaches intended to sense individual fake images. This subcategory emphases on deepfake video recognition approaches and classifies them into two clusters: approaches that uses chronological features and those that explore visual artifacts inside frames.

3.2.1 Temporal Features across Video Frames Based on the opinion that temporal coherence is not imposed efficiently in the procedure of deepfakes, Sabir et al. [10] use of spatio-temporal features of audiovisual streams to sense deepfakes. Video manipulates to carry out on a frame-by-frame basis so that low level objects formed by face manipulations are supposed to further clear themselves as temporal artifacts with irregularities among frames.

A recurrent convolutional model (RCN) was projected based on the combination of the convolutional
network Dense Net and the gated recurrent unit cells to exploit temporal inconsistencies across frames see Fig. 3. The proposed technique is verified on the Face Forensics++ data set, which contains 1,000 videos , and displays promising outcomes.
Guera and Delp [11] detailed that deepfake videos comprise intra-frame conflicts and temporal conflicts among frames. Then they proposed the temporal-aware pipeline technique that uses CNN and long short term memory (LSTM) to detect deepfake videos. CNN is employed to extract frame- level features, which are then fed into the LSTM to create a temporal sequence descriptor.
A fully-connected network is used afterwards for classifying doctored videos from real ones based on the sequence descriptor (Fig. 4).

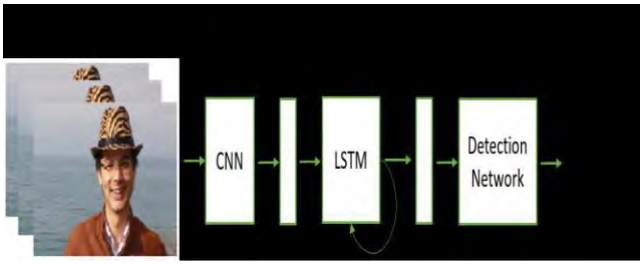


Fig. 4: A deepfake detection using CNN and LSTM.

Similarly, the use of a physical signal, eye blinking, to notice deepfakes was planned in [9] based on the opinion that a individual in deepfakes has a lot fewer frequent blinking than that in original video. A fit adult human would usually blink somewhere between 2 to 10 seconds, and every open and close the eye would take 0.1 to 0.4 seconds. Deepfake procedures, however, frequently use face images existing online for training, which usually demonstrate persons with open eyes. Therefore, without blinking images of persons, deepfake algorithms do not have the ability to produce fake faces that can blink typically. To distinguish actual and false videos, Li et al.
[9] initially decompose the videos into frames where face areas and then eye region are mined based on six eye signs. After some stages of pre-processing such as aligning faces, extraction and scaling the bounding boxes of eye sign points to generate new arrangements of frames, these cropped eye area arrangements are distributed into long-term recurrent convolutional networks (LRCN) for dynamic state prediction. The LRCN contains of a feature extractor built on CNN, a sequence learning built on long short-term memory (LSTM), and a state prediction built on a fully connected layer to forecast probability of eye open and close state. The eye blinking illustrates robust temporal dependencies and therefore the application of LSTM helps to capture these time-based patterns efficiently. The blinking rate is calculated built on the predicted outcomes where a blink is defined as a highest above the threshold of 0.5 with interval less than 7 frames. This technique is assessed on a dataset collected from the internet containing of 49 videos and their matching fake videos produced by the deepfake algorithms.

3.2.2 Visual Artifacts within Video Frame

As observed in the preceding subsection, the approaches using temporal patterns across video frames are typically based on deep recurrent network models to sense deepfake videos. This subsection inspects the additional

method that usually decomposes videos into frames and explores visual artifacts within single frames to get distinguished features. These features are then dispersed into either a deep or shallow classifier to discriminate between false and true videos. We thus assemble approaches in this subsection based on the categories of classifiers, i.e. either deep shallow.

3.2.2.1 Deep classifiers

Deepfake videos are usually formed with few resolutions, which need a face warping method (i.e., rotation and shear) to matches the original ones. Because of the resolution contradiction between the warped face part and the nearby context, this procedure leaves objects that can be noticed by CNN methods such as VGG16, ResNet50, ResNet101 and ResNet152. A deep learning technique to spot deepfakes based on the objects observed throughout the face warping step of the deepfake generation algorithms was proposed in [12]. The proposed structure is assessed on two deepfake datasets, i.e. the UADFV and Deepfake TIMIT. The UADFV dataset [15] contains 49 actual videos and 49 forged videos with 32752 frames in total. The Deepfake TIMIT dataset contains bad quality videos of size 64*64 and another good quality videos of 128 x 128 with 10537 unique images and 34,023 forged images mined from 320 videos for each set. Performance of this technique is associated with other predominant approaches such as the face tampering recognition technique two-stream NN, Head Pose [15], and two deepfake recognition MesoNet methods, i.e. Meso-4 and MesoInception-4 [13]. Advantage of the proposed technique is that it need not to produce deepfake videos as forged examples before training the detection methods. As an alternative, the false examples are produced dynamically by mining the face area of the unique image and aligning it into several scales before applying Gaussian blur to a scaled image of chance choice and warping back to the unique image. This decreases a huge time and computational resources associated to other approaches, which need deepfakes are produced in advance. Recently, Nguyen et al. [14] proposed the usage of capsule networks for distinguishing doctored images and videos. The deepfake face swap dataset created by Afchar et al. [13].

IV. RESEARCH ISSUES

Deepfakes are gradually damaging to confidentiality, society safety and democracy. Diverse approaches for detecting deepfakes are proposed by many researchers as soon as this menace was familiarized. Initial efforts were based on handcrafted features gained from objects and irregularities of the fake video synthesis procedure. Current methods, additionally, uses deep learning to automatically mine prominent and discriminative features to spot deepfakes. Using recognition approaches to spot deepfakes is critical but considering the actual intent of persons reproducing deepfakes is even more important. This needs the decision of users built on social context in which deepfake is revealed. This is serious as deepfakes are getting more photorealistic and it is extremely projected that recognition software will be lagging behind deepfake creation techniques.

4.1 Limitations of Existing system

In the technique where the recognition is done by eye blinking technique [9] the dataset used was very small thus system accuracy will be reduced. The experimental outcomes show great performance of the planned technique in detecting false videos, which can be further enhanced by seeing dynamic pattern of blinking. Detection is done by frame by frame basis so there might be chances of inconsistency [10]. RNN gives better result than CNN [11][12][13].

4.2 Effects of Deepfake

Misinformation: Folks are more probable to have a response to misinformation in the form of forged image, audio, and video content, which allows the doctored media to spread more rapidly than purely written fake data. Additional, images and video have been recommended to generate a Mandela effect, the formation of memories that never happened.

Exhaustion of serious thinking: It will take more effort for people to determine whether data is true, specially when it does not originate from reliable actors. Uncertainty around content reliability might also put off an distinct from sharing correct content, dropping the distribution of true information.

The deceiver's dividend: The presence of entirely artificial content offers an avenue for actors to deflect charges of impropriety based on footages and video, by appealing the source material has been faked.

These results are troubling and will be most persistent in the future, as deepfake excellence increases and social alertness lags

V. CONCLUSION

Deepfakes have started to wear down trust of people in media contents as seeing them is no longer corresponding with believing in them. This is critical nowadays as the technologies for creating deepfakes are progressively approachable and social media platforms can spread those fake contents rapidly. This paper has revised the state-of-the-art methods and a summary of typical approaches are provided. It is noticeable that a fight between those who use progressive machine learning to create deepfakes with those who make effort to detect deepfakes is rising. Deepfakes' good quality has been increasing and the

performance of discovery methods needs to be enhanced consequently. Detection methods are still in their early stage and various methods have been suggested and evaluated but using fragmented data sets. An approach to progress performance of detection methods is to create a rising updated standard data set of deepfakes to authorize the ongoing development of discovery methods. This will simplify the training process of discovery models, particularly those based on deep learning, which involves a large training set. The current detection methods mostly focus on disadvantages of the deepfake generation pipelines, i.e. finding weakness of the participants to attack them. This kind of information and knowledge is not always available in adversarial surroundings where attackers usually attempt not to reveal such deepfake creation technologies. This is a real challenge for detection method development and a future research needs to focus on introducing more robust, scalable and generalizable methods.

References

1)Tewari, A., Zollhoefer, M., Bernard, F., Garrido, P., Kim, H., Perez, P., and Theobalt, C. (2018). High- fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. DOI: 10.1109/TPAMI.2018.2876842.

2)Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X.,Huang, X., and Metaxas, D. N. (2019). StackGAN++: Realistic image synthesis with stacked generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8), 1947-1962.

3)Lyu, S. (2018, August 29). Detecting deepfake videos in the blink of an eye. Retrieved from http://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072

4)Damiani, J. (2019, September 3). A voice deepfake was used to scam a CEO out of $243,000. Retrieved fromhttps://www.forbes.com/sites/jessedamiani/2019/03 avoice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

5)Chorowski, J., Weiss, R. J., Bengio, S., and Oord, A. V.

D. (2019). Unsupervised speech representa- tion learning using wavenet autoencoders. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 27(12), pp. 2041-2053

6)Korshunov, P., and Marcel, S. (2019). Vulnerability assessment and detection of deepfake videos. In The 12th IAPR International Conference on Biometrics (ICB), pp.1-6.

7)Xuan, X., Peng, B., Dong, J., and Wang, W. (2019). On the generalization of GAN image forensics. arXiv preprint arXiv:1902.11153.

8)Agarwal, S., and Varshney, L. R. (2019). Limits of deepfake detection: A robust estimation viewpoint. arXiv preprint arXiv:1905.03493.

9) Li, Y., Chang, M. C., and Lyu, S. (2018, December). In ictu oculi: Exposing AI created fake videos by detecting eye blinking. In 2018 IEEE International Workshop on

5

Information Forensics and Security (WIFS) (pp. 1-7). IEEE.

10) Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., and Natarajan, P. (2019). Recurrent convolutional strategies for face manipulation detection in videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 80-87).

11) Guera, D., and Delp, E. J. (2018, November). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE

12) Li, Y., and Lyu, S. (2019). Exposing deepfake videos by detecting face warping artifacts. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 46-52).

13) Afchar, D., Nozick, V., Yamagishi, J., and Echizen, I. (2018, December). MesoNet: a compact facial video forgery detection network. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-7). IEEE

14) Nguyen, H. H., Yamagishi, J., and Echizen, I. (2019, May). Capsule-forensics: Using capsule networks to detect forged images and videos. In 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2307-2311). IEEE.

15) Yang, X., Li, Y., and Lyu, S. (2019, May). Exposing deep fakes using inconsistent head poses. In 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8261-8265). IEEE.

16) Matern, F., Riess, C., and Stamminger, MAIN. (2019, January). Exploiting visual artifacts to expose deepfakes and face manipulations. In 2019 IEEE Winter Applications of Computer Vision Workshops (ICCVW) (pp. 83-92). IEEE

17)Koopman, M., Rodriguez, A. M., and Geradts, Z. (2018). Detection of deepfake video manipulation. In The 20th Irish Machine Vision and Image Processing Conference (IMVIP) (pp. 133-136).

18)Hsu, C. C., Zhuang, Y. X., and Lee, C. Y. (2019). Deep fake image detection based on pairwise learning. Preprints, 2019050013.

19)Zhang, Y., Zheng, L., & Thing, V. L. L. (2017). Automated face swapping and its detection. 2017 IEEE 2nd International Conference on Signal and Image Processing(ICSIP).doi:10.1109/siprocess.2017.8124497

Comparative Analysis Of Clustering Algorithms

Harshali Desai, Anand Khandare
Thakur College of Engineering and Technology
University of Mumbai
harshalidesai28@gmail.com, anand.khandare1983@gmail.com

Abstract- **Clustering is an unsupervised learning technique in which similar datapoints are grouped together into groups called clusters. Each cluster contain datapoints that are similar to each other and dissimilar to datapoints in other groups. Clustering is achieved by various algorithms like K-means, DBSCAN, PAM, Hierarchical clustering etc. Every algorithm has its own drawbacks which has to be overcome. The most common drawback of algorithms is that they do not work well for large datasets. Accuracy of algorithms is degraded for large datasets. Also, this type of dataset increases processing time of algorithm. This paper compared different clustering algorithms on various standard datasets and analyzed their performance with respect to various performance parameters such as time complexity, accuracy and various cluster quality measures. From this analysis it is found that there is a scope to enhance the performance of these algorithms.**

Keywords- clustering, quality measures, performance parameters,accuracy

I. LITERATURE REVIEW

Clustering algorithm uses distance metric or similarity metric in order to compute similarity between the datapoints. Choosing any appropriate clustering algorithm totally depends on the type of application or datasets we are working with. It is very easy and efficient to work with numerical dataset. But when it comes with working with the other datasets like categorical, time series data it becomes very difficult to handle problems with these datasets.[1]

When a cluster belongs to an only single and distinct cluster then such clustering is called hard clustering, whereas a cluster belonging to different cluster is called soft clustering.

The author(s) have highlighted various merits and demerits of K-means, KNN clustering algorithm. Number of the clusters to be formed is the major factor while selecting correct clustering algorithm.[2]

Clustering is also referred as cluster analysis. The author aims to give basic understanding of clustering by explaining the type of clusters and various clustering methods. In well separated clusters the dissimilar datapoints lies far away from each other.[3] In clustering technique large dataset is divided into small groups called as clusters. In other words, clusters are group of similar data. These clusters have following properties:

- Each cluster is homogeneous, which states that the datapoints in same cluster are similar to each other.
- Every cluster must be different from every other clusters, which states that the datapoints in different clusters are dissimilar to each other.[4]

K-Means algorithm is one of the simplest and most widely used partitioning based clustering algorithm. The goal of K-mean algorithm is to partition datapoints into ‘k’ clusters based on their similarities. The similarity between the datapoints is calculated with help the of random or mean value called cluster centroid. In most of the cases Euclidean distance is used as dissimilarity measures for k-means algorithm. Calculating the similarity between clusters is dependent on the implementation of the model. All the clustering algorithm has its own measures to calculate the similarity or dissimilarity.

The major drawback of K-Means is prediction of accurate ‘k’ value, as different ‘k’ gives different clusters. DBSCAN algorithm is used to discover clusters of arbitrary shape, unlike K-means which discovers spherical clusters. Also, DBSCAN algorithm does not require number of clusters as initial parameter.[5]

Weka tool is an open source software which is a collection of machine learning algorithms mainly used for data mining tasks. The author(s) has compared the clustering algorithms namely K-means, Density based algorithm, Hierarchical algorithm using Weka tool in terms of efficiency and accuracy. K-Means algorithm yields more accuracy and consumes less time with comparison to Hierarchical, Density clustering algorithm.[6]

II. INTRODUCTION

In machine learning, the learning is divided into mainly 3 categories- Supervised, Unsupervised and Reinforcement. In supervised the machine or model will be trained using labelled data (i.e. data

which is already tagged with correct output). But in case of unsupervised learning the model is not trained usinglabelled data. The model in such scenario have to find the hidden patterns or structures or groups from input data based on similarity. One such technique of unsupervised learning is clustering.

Clustering is a process of dividing the datapoints or population into groups of similar datapoints. So, the output of clustering will be groups having datapoints which are similar to the datapoints in the same group and are dissimilar to the datapoints of other groups.

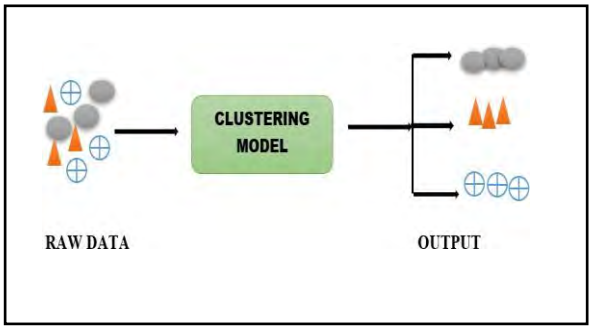


Figure 1. Overview of Clustering process.

The Figure 1. depicts the working of clustering process. In above figure, the raw data is passed as an input to the clustering algorithm or model and the resultant clusters are formed based on the similarity between the objects or datapoints.

There are various clustering algorithms. In this paper, we have studied and analysed five most common clustering algorithms
– Hierarchical, K-means, PAM, Mini batch K-means, DBSCAN.

1. K-Means

K-means algorithm is a type of unsupervised partitioning based clustering algorithm. The goal of this algorithm is to divide the datapoint into 'k' clusters using centroid. The k-means clustering algorithm assigns datapoints to the cluster by finding the minimum distance between the datapoints and centroids. It then iterates through this technique in order to perform more accurate clustering over a time.

In this method the minimum distance between datapoint and centroid are calculated by Euclidean distance formula.

Euclidean distance formula is given by:

$$\text{Distance}(p, q) = \text{Sqrt}\left(\sum (qi - pi)2\right).$$

Algorithm for K-means is as follows:

Step 1. Randomly Select 'k' centroids C1 to Cn.
Step 2. Calculate distance between each data point and centroid using Euclidean distance formula.
Step 3. Assign datapoint to the respective cluster having minimum distance.
Step 4. Recalculate the new centroids.
Step 5. Repeat Step (2) to (4) till same clusters are obtained.

After applying K-means on iris dataset for sepal length and sepal width the resultant three clusters formed are represented below and the black dot represents the centroids.

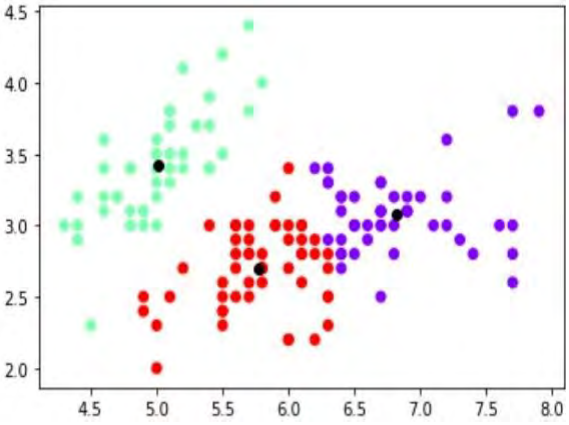


Figure 2. K-Means clusters for Iris dataset

Advantages:

1) When centroids are recalculated object can change the cluster i.e. can move into another cluster.
2) Only work well for spherical clusters.
3) It is very easily implemented for numeric datasets.
4) Tighter clusters are produced.

Disadvantages:

1) Difficult to predict initial 'k' value i.e. number of clusters.
2) Random selection of centroids can also yield less accurate results.
3) Does not work well for categorical values.
4) This algorithm can discover spherical clusters. Thus, it is prone to outliners.

2. HIERARCHICAL CLUSTERING

This technique is also referred as hierarchical cluster analysis (HCA). Similar clusters are merged or dissimilar clusters are splitted depending on the type of hierarchical algorithm used. This technique is broadly classified as – Agglomerative and Divisive Hierarchical clustering.

Agglomerative Hierarchical clustering - It starts

with treating each datapoint as an individual cluster and then go on merging similar clusters to form a final clusters. Therefore, this method corresponds to bottom-up approach.

Divisive Hierarchical clustering- All the datapoints are consider as a single cluster and then in each iteration, we

separate datapoints from the cluster which are dissimilar. Therefore, this method corresponds to top-down approach.

The following linkage criterion are used which determines what distance to use between sets of datapoints. This algorithm then merges the clusters that minimize this criterion.

- Single Linkage: In this method, the distance between two clusters is defined as the shortest distance between two points in each cluster.
- Complete Linkage: The distance between two clusters is defined as the longest distance between two points in each cluster.
- Average Linkage: In average linkage hierarchical clustering, the distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster
- Ward Linkage: It minimizes the variance of the clusters being merged.

The agglomerative clustering is performed as follows:

Step 1. Consider each datapoint as a separate cluster. Given a dataset (d1, d2, d3…dn) of size 'n', treat each datapoint in dataset as a separate cluster (c1, c2, c3… cn).
Step 2. Compute the distance matrix.
Step 3. Then perform below steps until similar clusters are merged.
   a. Identify the two clusters that are closest to each other.
   b. Merge the two similar clusters and update the distance matrix.

Hierarchical clustering is implemented on shopping dataset where the attributes for forming clusters are annual income and spending score.

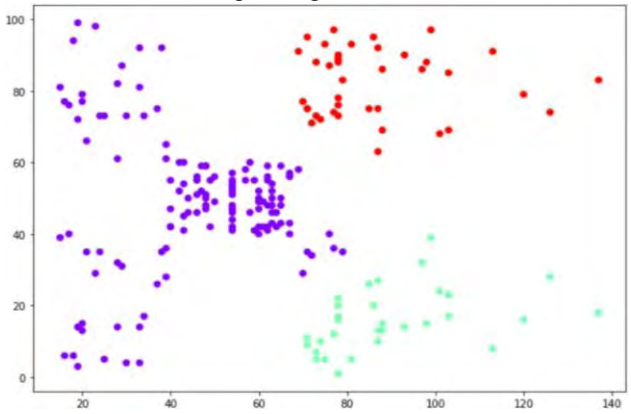


Figure 3. Hierarchical clustering on Shopping dataset

Advantages:

1) No priori information about the number of clusters required.

Disadvantages:

1) Not suitable for huge dataset due to high time complexity.
2) Loose clusters are formed as compared to K-Means.
3) Algorithm cannot undo previous step once merging or splitting is done.

3. DBSCAN

Partitioning and hierarchical clustering are suitable for only compact, spherical and well separated clusters. Moreover, they are also severely affected by the presence of noise and outliners in the dataset. This drawback is overcome by DBSCAN, which discovers arbitrary shapes based on density.

The idea behind this algorithm is that for each point of a cluster its radius has to contain at least minimum number of points. In simple words, DBSCAN works by identifying dense region, which are measured by the number of datapoints close to a given point.

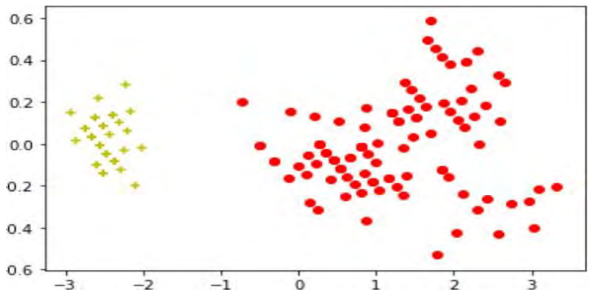
The algorithm works on two fundamentals:

- eps (epsilon): Radius of neighbourhood around the point.
- MinPts: It determines minimum number of neighbours within the 'eps.

DBSCAN algorithm is as follows:

Step 1. Pick random point 'p' and mark is as visited.
Step 2. Get all the points which falls in neighbourhood (upto eps distance) of point 'p' and store it in set S.
Step 3. If S >= minPts, then
   a. Consider point 'p' as first point of a new cluster.
   b. Consider all points (which are member of set S) as other points in this cluster.
   c. Repeat the above step b for all points of S.
Step 4. Else mark p as outliner or noise.
Step 5. Repeat step 1-5 till clustering is completed and all points are visited.

Two clusters are formed when DBSCAN is applied on iris dataset.



Advantages:

1) Does not require to specify the number of clusters to be generated.
2) It can discover arbitrary sized and arbitrary shape clusters very well.
3) Outliers or noise are easily identified.

Disadvantages:

1) Doesn't perform well when clusters are of varying densities.
2) Selection of DBSCAN parameters (eps and MinPts) is very tricky.

4. K-Medoid

K-Medoid also called as Partitioning Around Medoid (PAM) was proposed by Kaufman and Rousseeuw. A medoid is a point in a cluster which has minimum dissimilarity with all other points in the cluster. Clusters are formed based on the randomly selected medoids from dataset. Then the other non-medoid datapoints are grouped with that medoid to which it is most similar.

The basic algorithm for K-medoid is:

Step 1. Randomly select k from 'n' datapoints as the medoids.
Step 2. Assign each datapoint to the closest medoid.
Step 3. For every medoid 'm' and every data point which is associated to m swap m and o and compute the total cost (i.e. the average dissimilarity of a datapoint to all the data points associated to m). Then select the medoid which has lowest cost.
Step 4. Repeat the steps 2 and 3 until there same clusters are obtained.

Advantages:

1) Less sensitive to outliner compared to K-Means.

Disadvantages:

1) Not suitable for clustering arbitrary shape clusters.
2) It is expensive than K-Means, as it compares each medoid with entire dataset in every iteration.

5. Mini Batch K-Means

It is similar to K-means with only difference that here the computation is only done on random batch of observation.
This algorithm works by taking small random batches of fixed sized data. This is done so as to reduce the memory storage. Therefore, this approach significantly reduces the total time required by the algorithm to fit the data.

Mini Batch K-Means algorithm is as follows:

Step 1. Random samples(batches) are chosen from the dataset.
Step 2. Assign these batches to nearest centroids.
Step 3. Update the centroids.
Step 4. Perform above Steps until it reaches pre- determined number of iterations.

Advantages:

1) Reduces the computational time as compared to K- Means.

Disadvantages:

1) Still requires random selection of centroids.
2) Need to specify number of clusters in advance.

III. CLUSTERING PERFORMANCE MEASUREMENT PARAMETERS

Evaluating the clustering algorithm is helpful in determining how accurate the algorithm is running. Evaluating a clustering model is not very easy therefore many evaluation parameters has been proposed which are used to evaluate the clustering algorithms.
Some of these evaluation parameters are discussed with an example:

1. Within Sum Square Error (SSE)

It is defined as the sum of the squared differences between each datapoint and cluster centroid within cluster.

Formula is given by,
Within SSE = $\sum (Ci - Xi)^2$.

2. Between Sum Square Error (SSE)

It is defined as the sum of squared differences between two clusters.

It is given by formula,
Between SSE = $\sum$ (Global Mean - Ci)^2 * number of datapoint in the cluster.

3. Total Sum Square Error (SSE)

It is given by formula,
Total SSE = Total Within SSE + Between SSE

4. Accuracy

It defined as a measure to determine the quality of

clustering. It is calculated as,
Accuracy = (Between SSE/Total SSE) * 100

5. Time Complexity

Time taken by clustering algorithm to form k specified clusters. Time complexity of a clustering algorithm depends upto the number of clusters, number of iterations etc.

For below K-Means example we have calculate the above performance parameters.

Example 1: Consider the dataset {2,3,10,12,15,21,22,23,24} and number of clusters (k) = 3.

Randomly select initial centroids: C1 = 2, C2 = 15, C3 = 23

Calculate the Euclidean distance between each datapoint and centroid and assign the datapoint to cluster having minimum distance.

Table 1. K-Means 1st iteration table

Data Point (X)	Dist|C1-X|	Dist|C2-X|	Dist|C3-X|	Cluster
2	**0**	13	21	Cluster1
3	**1**	12	20	Cluster1
10	8	**5**	13	Cluster2
12	10	**3**	11	Cluster2
15	13	**0**	8	Cluster2
21	19	6	**2**	Cluster3
22	20	7	**1**	Cluster3
23	21	8	**0**	Cluster3
24	22	9	**1**	Cluster3

After calculating Euclidean distance between each datapoint and centroid and assigning the datapoint to cluster having minimum distance the clusters formed are:

Cluster1= {2, 3}

Cluster2 = {10, 12, 15}

Cluster3 = {21, 22, 23, 24}

Now, recalculate the centroid by taking mean of the clusters. Therefore, the new centroids are:

C1` = (2 + 3) / 2 = 2.5
C2`= (15 + 12 + 10) / 3 = 12.33
C3` = (21 + 22 + 23 + 24) / 4 = 22.5

Again, calculate the Euclidean distance between each datapoint and centroid and assigning the datapoint to cluster having minimum distance.

Table 2. K-Means 2nd iteration table

Data Point (X)	Dist|C1`-X|	Dist|C2`-X|	Dist|C3`-X|	Cluster
2	0.5	10.33	20.5	Cluster1
3	0.5	9.33	19.5	Cluster1
10	7.5	2.33	12.5	Cluster2
12	9.5	0.33	10.5	Cluster2
15	12.5	2.67	7.5	Cluster2
21	18.5	8.67	1.5	Cluster3
22	19.5	9.67	0.5	Cluster3
23	20.5	10.67	0.5	Cluster3
24	21.5	11.67	1.5	Cluster3

Therefore, the clusters formed are:

Cluster1= {2, 3}

Cluster2 = {10, 12, 15}

Cluster3 = {21, 22, 23, 24}

Since, the clusters obtained after recalculating the centroid are same, we can say than the clusters formed are correct and final and then we can terminate the algorithm.

For above example (Example 1) the performance parameters are defined and calculated as:

1. Within Sum square error (SSE)

Within SSE for Cluster1= $(2.5-2)^2 + (2.5-3)^2$ = 0.25 Within SSE for Cluster2= $(12.33-10)^2$ + $(12.33-12)^2$ + $(12.33-15)^2$ = 12.66
Within SSE for Cluster3= $(22.5-21)^2 + (22.5-22)^2+(22.5-23)^2+(22.5-24)^2$ = 5

Total Within SSE = Within SSE for Cluster1+ Within SSE for Cluster2+ Within SSE for Cluster3.

Therefore, Total Within SSE = 0.5 + 12.66 + 5 = 18.16

2. Between Sum square error (SSE)

Global Mean = (2+3+10+12+15+21+22+23+24) / 9 = 14.66

Therefore,

Between SSE = $[(14.66-2.5)^2 * 2 + (14.66-12.33)^2 * 3 + (14.66 -22.5)^2 * 4] =$ **557.8803**

3. Total Sum Square Error (SSE)

Total SSE = Total Within SSE + Between SSE
= 18.16 + 557.8803 = 576.0403

4. Accuracy

It is calculated as,
Accuracy = (Between SSE/Total SSE) * 100

Therefore,
Accuracy = (557.8803 / 576.0403) * 100 = **96.84**

IV. CONCLUSION

Clustering is a technique of dividing the entire population into small groups. There are many clustering algorithms, but each algorithm faces some drawbacks. These algorithms are analysed using various clustering evaluation parameters. There are many ways to determine the quality of clusters. Large number of dataset degrades the quality of clusters and also slow downs the performance of the clustering algorithm.

V. References

[1] Review based on data clustering algorithms, Proceedings of 2013 IEEE International Conference on Information and Communication Technologies (ICT 2013)

[2] A Comparative Study of Data Clustering Algorithms, International Journal of Computer Applications (0975 – 8887) Volume 83 – No 15, December 2013

[3] Survey of Cluster Analysis and its Various Aspects, IJCSMC, Vol. 4, Issue. 10, October 2015, pg.353 – 363

[4] A Survey of Clustering Techniques, International Journal of Computer Applications (0975 – 8887) Volume 7– No.12, October 2010

[5] Comparative study of data clustering techniques, 2016 IJEDR | Volume 4, Issue 4 | ISSN: 2321-9939

[6] A Comparative Analysis of Clustering Algorithms, International Journal of Computer Applications (0975 – 8887) Volume 100 – No.15, August 2014

Intelligent Traffic Monitoring System

Ms. Vidhi Punjabi , Mr. Akash Singh , Ms. Urvi Shukla
Department of Computer Engineering
Thakur College of Engineering And Technology
Mumbai, India
vidhpunjabi@gmail.com , akash.s.13201@gmail.com , urvishukla20@gmail.com

Abstract: **The increasing population around the world, development , growth and urbanization and also the increase in the standard of living has lead to the increase vehicular usage by the people since the people are moving more towards the private vehicles than the use of public transport vehicles. Due to this our traffic system is facing load of the increased vehicular traffic and it is leading to the crowded roads and cities and our traditional traffic control system is clearly inefficient in proper traffic control and management. However few such model or design exists which actually consider the flow of the vehicle at the node and consider the factors which include vehicle deviation, vehicle density, number of emergency vehicles present ,number of pedestrians crossing and also accidents. Here in this article we have reviewed the various systems available and used around the world and also tried to find out the possible improvements which are needed and tried to develop an algorithm which focuses and uses machine learning strategy to provide solutions which can be implemented in order to achieve the necessary needed improvements.**

Keywords: Intelligent Traffic System, Traffic light, Congestion, Vehicle Density.

I. INTRODUCTION

The ever-increasing population and urban development has lead to the problem of road congestion and heavy traffic jams and hence leading to various problems concerned with the environment and economy. Emergence of such scenario make it difficult for the traditional traffic monitoring system to provide efficient management of traffic . Also since the traffic is a matter of critical concern as it includes various constraints like vehicular density, number of pedestrians, number of emergency vehicles on the road, vehicle deviation and other during finding a approach for traffic management. The traffic light system used are also just designed in a way to show a particular signal light for a specific duration of time. Though there exists other systems through which we can change the duration for a traffic light based on the situation like peak hours and free hours. However they can't themselves take into account the factors of density of traffic on the road. All together this traffic congestion problems result in unexpected delays and makes it difficult to reach the required destination on time especially in case of emergency services like ambulance, fire brigade vehicles. Hence taking into consideration there was need of the intelligent monitoring system which must incorporate the algorithms and techniques which should be able to meet these requirements. Hence here we study the various algorithms and technologies so developed for better and effective traffic management and monitoring.

II. LITERATURE SURVEY

[1] This system focuses on Congestion Control, Emergency Vehicle Clearance and also Stolen Vehicle Detection. The system would be designed in a way such that it will be detecting the number of vehicles entering into the system using Infrared Sensors (IR Sensors). These IR Sensors are connected to a microcontroller. The System proposes that this microcontroller would be able to detect the vehicle density and will be automatically able to control traffic. This is done by controlling the traffic signal lights in a four road junction. Here for the experiment purpose suggests the use of RFID tags with a RFID reader with frequency 125kHz. RFID tag of vehicle comes in the range of the receiver will transmit that unique RFID to the reader. This RFID reader is connected to the microcontroller which counts the RFID tags. This RFID tags and reader work using IR sensors. Also for Stolen vehicle Detection the stolen vehicle RFIDs will be kept in a list and this vehicle will come in a contact of a RFID reader and it will compare the tag with the list of tags present in the list and it will detect the stolen vehicle. For Emergency vehicle clearance system working will be done with the combination of

ZigBee transmitter and receiver. The vehicles which are used for emergency purposes consists of ZigBee transmitter which will be switched on during the emergency use. The signal transmitted by such vehicles will be detected by ZigBee receiver. Detecting such signal will automatically make the system to give preference to that lane consisting of emergency vehicle.

[2] This system basically focuses on Vehicle Detection and recognition for Traffic Surveillance System. It manages vehicle detection and its recognition on the basis of static images using image processing. The images are read by a classifier which has functions based on the machine learning algorithms such as Ada Boost algorithm and Haar .This allows recognition of the region of interest. Further the additional training is carried out for analysing the different types of vehicle samples.

[3] This proposed system basically focuses on customizing the duration of green signal based on the current traffic conditions and it aims of this system is to determine the best possible cycle duration and best possible phase depending upon the current situation i.e. the incoming traffic from each of the four lanes .It also make sure that duration of each phase should be flexible so that it must not result into traffic congestion in any of the remaining lane. The vehicle density of each road is calculated and depending upon the results traffic signals will be programmed such that more crowded road will get green signal for greater duration of time.

III. PROPOSED SYSTEM

The main disadvantage of the traditional system is that it doesn't understand important factors like peak hours of traffic or sudden factors like a procession on the road. Hence, there are times when vehicles have to wait for too long before the signal releases. In worse situations, a platoon (a term coined by traffic engineers), which is a group of vehicles cannot make it all through the traffic without any kind of interruptions, thereby, a particular platoon breaks into two and the latter half of the platoon might have to stay on the same signal twice. The proposed system ensures that no particular platoon breaks. It also ensures that there are no pileups during peak hours. Moreover, this system also ensures that no drivers are induced to jump signals during lean hours.

IV. THE DETECTOR

This system requires the installation of an economically-friendly detector, which will measure the maximum traffic upto a specifically calculated distance.

The detector is a wideband spread-spectrum radar. It is very inexpensive and it has extremely accurate range discrimination. It can also penetrate many types of materials, including concrete. It has a range of about 20 feet, so it may be useful as an inexpensive, single-lane vehicle detector. It is predicted that the sensor, when made in production quantities, would cost much less than $10 per sensor. Because of their accurate range discrimination, they have a very well defined field of detection. They could become a cheap alternative to magnetometer probes. Their ability to detect range provides additional information for future traffic control systems.

We will develop a broadband transmitter/receiver pair to be used with these sensors. This would eliminate the need for communication lines between the sensor and the controller.

V. METHODOLOGY

Our model works on the simple logic of "freeing the congestion first" during peak hours. And "freeing the easy transit" during lean hours. If the situation it's a mix of both then, our algorithm alternatively executed its task.This method defines traffic as still movement of vehicles for 200 meters of a 4-lane passage. The mentioned range was chosen after a careful consideration of traffic in different locations of the city of Mumbai.

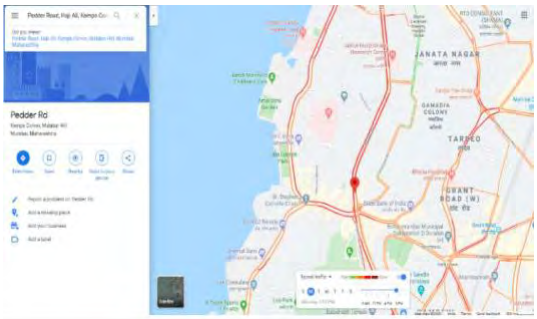


Fig.1: Showing the traffic of one of the busiest roads of Mumbai-Peddar Road.

In the above diagram, it is very clearly depicted that how all the roads are jam packed. The shockwave theory

The second consideration of this method is that each row of vehicle in a traffic is taken as 5 metres. This calculation is an average of a vehicle as small as a rickshaw and a vehicle as large as a bus. With the help of the spread-spectrum wideband radar, the system would analyse the situation of the traffic and categorize it in either of the three option viz., i] High ii] Low iii] Optimal.

For this analysis the setup would require a maximum of 40 hops from the start of the signal to 200 metres away from it. This calculation is given by:

Thus, the time complexity of our counter would be O(n). Several other factors are taken into consideration, such as the width of the road. Therefore, the road is divided into particularly three halves, i]4 laned passage ii]3 laned passage iii] 2 laned passage(due to the barriers of the roads such as sudden accidents leading to road rage, procession on the road or coherent construction of roads). In all the above scenarios, the waiting time and the release time would be different. These different values would be already pre-set by the traffic engineers based on the local factors and environment of a particular intersection.

Another consideration proposed by this approach is the modification based on the movement of the traffic. At each approach to the intersection there are three direction the vehicles can go movements. These movements are namely, left, through and right. Right and through are typically grouped as a single movement. So an intersection has 8 vehicular and 4 pedestrian movements.

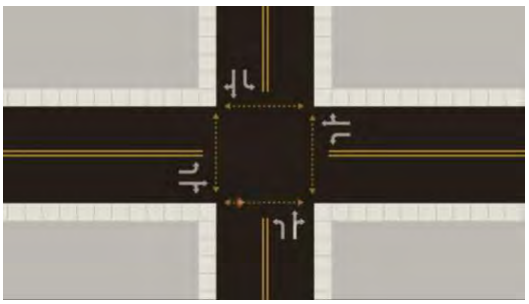


Fig.2 The diagram depicts an 8 way vehicular movement

In such a case, the movements can be phased into phases. For example, right turn movements of opposite intersection can be grouped into a single phase. Because they both can move at the same time without any kinds of interruptions or conflicts.Such a method decreases the overall waiting time of the particular intersection. Because instead of waiting for three other signals before your turn come, one only has to wait for 1 signal. However, the disadvantage of such a method is that the movement gets restricted. As the vehicles get prohibited from taking one of the turns.

The second situation is when no turns are restricted and one path has to wait for the other three to complete their cycles before releasing its traffic again.

The fifth consideration is to decrease the noise levels of the road. Our model not only aims at controlling the flow of the traffic but it aims to minimize congestion. Hence, there is a need to control the average decibels on the road.

For the same reason,

A sensor will be installed which will sense the noise . It basically work as when the red light will come to an end and will shortly turn into green light and if before its end if noise (horn) will be made by the travellers and if that noise level with cross the set value the red light automatically will increase by 90 sec or more. This mechanism will help to control the noise level , for smooth functioning of the system and many more..

VI. WORKING

To explain the working our our proposed solution we have take the following considerations:

1. The pivot is considered to be the south-most intersection. This means that the default scanning of the traffic will start from the above intersection. And then go in the anti-clockwise direction.

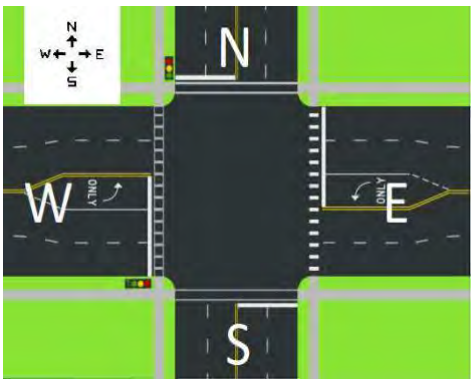


Fig. 3 depicts the naming of the direction of road**s**

2. The default time of the signal will be taken as 35s, 70s, 135s and 0s (0s is taken because in a crossroad three roads would be waiting while only one can be released)
This time will not be displayed on the screen. The purpose of this time is to act as a default after the which the four signal will have to mandatorily swap.
3. 200 meters is the situation with heavy traffic while 20 meters is the situation with light traffic

CASE 0: OPTIMAL TRAFFIC

This scenario would occur when then the jam is less than 200 meters but more than 20 meters. Such a system would word similar to the traditional system Since the pivot is decided to be south, the start would be considered from south and then the traffic will move in this way:

	Pass 1	Pass 2	Pass 3	Pass 4
South	35s	0s	135s	70s
West	70s	35s	0s	135s
North	135s	70s	35s	0s
East	0s	135s	70s	35s

Fig. 4 depicts the waiting time of each road in each particular pass.
Signal 1 gets free, after waiting for 35 seconds. It is released for 32 seconds and the buffer time is 3 seconds.
After (35+32+3) seconds, considered from time 0, the second signal will get released for 32 seconds and the buffer time is 3 seconds.
After (35+35+32+3) seconds, considered from time 0, the third signal will get released for 32 seconds and the buffer time is 3 seconds.
After (35+35+35+32+3) seconds, considered from time 0, the fourth signal will get released for 32 seconds and the buffer time is 3 seconds.

VII. CASE 1: HEAVY TRAFFIC

PART A: ONLY ON THE SOUTH ROAD

In this situation, the timer which is supposed to be 35s as default automatically, becomes a countdown of just 10 seconds. Such a situation occurs when the other three roads have optimal traffic.
Signal 1 gets free, after waiting for 5 seconds. It is released for 36 seconds and the buffer time is 2 seconds.
After (5+36+2=43) seconds, considered from time 0, the second signal will get released for 32 seconds and the buffer time is 3 seconds.
After (35+43) seconds, considered from time 0, the third signal will get released for 32 seconds and the buffer time is 3 seconds.
After (35+35+43) seconds, considered from time 0, the fourth signal will get released for 32 seconds and the buffer time is 3 seconds.

In this situation, the release time is considered to be more for the region with high traffic as compared to the region of low traffic. It is calculated by estimating an average speed of a car in traffic as 10 kmph. Then, Distance/Speed=Time, which is calculated to be 36 sec. The above consideration is an important factor, as our goal is to make sure that a platoon doesn't break into 2 equal or unequal halves.

After the 4th signal is released of 1 cycle, the time calculated turns out to be 113s. While with the traditional method it was 135 seconds. Hence, 22 seconds per cycle is saved this way.

PART B: ON ANY OTHER ROAD

This situation is a simple alteration of Part a. If any other road, except south, have heavy traffic, then the pivot is temporarily changed to that particular road and then it proceeds in the same anti-clockwise manner.

CASE 2: HEAVY TRAFFIC ON MORE THAN 1 ROADS

Case 2 is nothing but a combination of part a and part b of case 1. In the circumstance, the priority would be given to the hroads with heavy traffic first. For example. if West and East have heavy traffic, then the flow of the traffic as per the anti-clockwise rule (with the pivot being West) would be: West East South North.
Note that the above algorithm was designed keeping in mind the priority of the flow of traffic.

CASE 3: LIGHT TRAFFIC

PART A: ONLY ON 1 OF THE CROSSROADS.

If the jam is under 20 meters from the traffic light then it is considered as light traffic. Light traffic is as dangerous as heavy traffics are because usually the riders are tempted to jump the signal due to the easy accessibility and also due to impatience.

Hence, the waiting time would be 5 seconds and the release time would be 12 seconds.

Signal 1 gets free, after waiting for 5 seconds. It is released for 10 seconds and the buffer time is 2 seconds.

After (5+10+2=17) seconds, considered from time 0, the second signal will get released for 32 seconds and the buffer time is 3 seconds.

After (35+17) seconds, considered from time 0, the third signal will get released for 32 seconds and the buffer time is 3 seconds.

After (35+35+17) seconds, considered from time 0, the fourth signal will get released for 32 seconds and the buffer time is 3 seconds.

After the 4th signal is released of 1 cycle, the time calculated turns out to be 87 seconds. While with the traditional method it was 135 seconds. Hence, 48 seconds per cycle is saved this way.

PART B: ON MORE THAN 1 ROADS

In this case, if there are two crossroads with light traffic and the 2 others with heavy traffic then alternatively light and then heavy traffic would be released. For instance: if South and West have heavy traffic while North and East have Light traffic then, the pivot would be North and the flow would be as follows:

North South East West

The above flow ensures the alternate release of light and heavy traffic turn-by-turn.

ALGORITHM

For making the algorithm, a datatype called timecount is used. timecount allows the program to increment or decrement the value of an variable by 1 every second.

Function car:

1. boolean p
2. count=5 meters
3. while (count =200)
4. if (p=true)
5. Count+=5
6. Return count.

The above function will run in correspondence to the wideband spread-spectrum radar and return how much length of the road is occupied to another function HLO.

Function HLO:

1. If count=200 Return h
2. If count<=20 Return l
3. If count<200 but count >=20 Return 0

This function categorizes the traffic different roads of an intersection into either of the three parts viz., H, L or O. H stands for high, L stands for low and O stands for optimal.

Main Function:

1. Storing values of H, L or O for 4 roads viz. in an array check [s, n, e, w]. Generate three arrays of datatype timecount namely, green, red and yellow
2. for i=1 to 4 if (i=4) i=0

if(check[i]==check[i+1])then

red[i]=red[i]+35 red[4]=0
if(red[i]=0)then yellow[i]=3

if(yellow[i]=0) then green[i]=32

if (green[i]=0) then restart the timecount variable
else if(check[i]='h')

red[i]=5
if(red[i]=0) then
yellow[i]=2

if (yellow[i]=0) then
green[i]=36

if(green=0) then
restart the timecount variable

else(check[i]=='l')
red[i]=5
if(red[i]=0) then
yellow[i]=2

if (yellow[i]=0) then
green[i]=10

if(green=0) then
restart the timecount variable

VIII. CONCLUSION

The main motive of this intelligent traffic model is to reduce the level of not only traffic congestion but also noise levels in a smarter way. It is a process which analyses the amount of traffic congestion at any given point of time and changes its methodology of working based on different situations. Since, the system is not

static or standardized a buffer period of 5 second is always present so the traveller gets alerted before the signal turns green or red as the case may be.

IX. FUTURE SCOPE

In dense urban areas, traffic congestion is often very self-limiting. It is estimated, that during peak times for every one person on the road there are many others at their home or work waiting for the congestion to clear up before heading out. This theory therefore predicts that any free up of traffic will be quickly filled up with more traffic bringing the congestion back to the same level from where it began. Hence, we can definitely, consider this factor in future to decrease congestion levels.

References

[1] Professor Sunil Ghane,Vikram Patel,Kumarasen Mudliar,Abhishek Naik. "Using AI and Machine Learning Techniques for Traffic Signal Control Management-Review." International Journal of Engineering Research & Technology(IJERT). Vol.6 Issue 11,Nov- 2017.

[2] Yong Tang ,Congzhe Zhang ,Renshu Gu ,Peng Li , Bin Yang, "Vehicle detection and recognition for intelligent traffic surveillance system," Springer Science + Business Media New York 2015.

[3] Sanchal Ramteke, Professor B.B. Gite."AI Based Traffic Signal Control System." International Research Journal of Engineering and Technology(IRJET).Vol 06.Issue 11,Nov-2019

[4] Vikas Gayah, Guajie Zheng, Hua Wei, Zhenhui Li."A Survey on Traffic Signal ControlMethods."https://www.researchgate.net/publication/332494086. Vol.1

Recent Trends in Soft Computing: A Case Study

Anam Khan, Dr. Megharani Patil
Department of Computer Engineering
Thakur College of Engineering & Technology
University Of Mumbai
Mumbai, India
anam.mail4u@gmail.com,megharani.patil@thakureducation.org

Abstract— ***Soft Computing is a branch of science and engineering that allows us for reasoning, thinking, analyzing, and detecting, and correlating the real-world problems to the biologically inspired methods. Soft Computing is the big motivation behind the idea of machines working on conceptual intelligence. As such, it's an extension of heuristics for solving complex problems that are too difficult to model mathematically. Soft Computing is tolerant of imprecision, uncertainty, and approximation and is different from hard computing. Soft Computing has a collection of techniques like ANN, Evolutionary computing, Fuzzy Logic, and Statistics. Though they are advantageous and separately applied techniques but when hybridized, solve complex problems very easily. Also, Big Data is paving its way to enter into the area of Soft Computing and build more intelligent and computationally accurate systems. Various soft computing techniques and emerging fields of soft computing where they are successfully applied are highlighted in this paper.***

Keywords— Soft Computing, Fuzzy Logic, Artificial Neural Network, Genetic Algorithm.

I. INTRODUCTION

Engineering is a task of solving complex real-world problems. For obtaining the low cost and feasible solution to these problems, numerous tools and technology are already available. Among them, Soft Computing is one of the powerful problem-solving technique for the problems which have no known method to find an exact solution.

The idea of soft computing came into light in the year 1981 when professor Lotfi A. Zadeh published his first paper on soft data analysis titled "What is Soft Computing". Professor Zadeh defined soft computing as "it is an innovative approach to design a system which parallels the extraordinary ability of the human mind and is computationally intelligent to learn and reason in an environment of imprecision and uncertainty". It is a multi- disciplinary system that is a combination of the fields of Fuzzy Logic, Artificial Neural Networks, Genetic Algorithms, Probabilistic Reasoning, and Neuro Computing. Soft Computing is the fusion of methodologies effectively designed and modeled to provide a solution to the real word problems, which are however too difficult to model mathematically. The primary goal of soft computing is to follow and match the human mind and its reasoning and thinking behavior as much as possible. It supports low solution cost and better rapport with reality. The final aim is to develop a computer or a machine which will work exactly similar to human-being thereby replicating the wisdom of human being into the computers in an artificial manner.

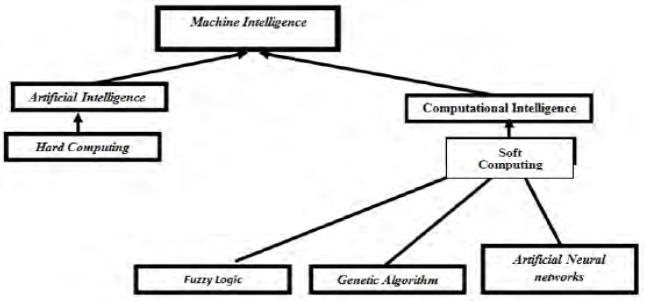


Figure 1: An Overview of Branches of Machine Intelligence

Soft Computing is probably going to play an especially important role in science and engineering, and eventually, its influence is extending farther. Building human-centered systems is an important task for scientists and engineers within the new millennium.

Fuzzy Logic (FL)

Fuzzy Logic is predicted on pure mathematics and provides methods for modeling and reasoning under uncertainty, a characteristic present in many problems, which makes FL a valuable approach. It allows data to be represented in intuitive linguistic categories instead of using precise(crisp) numbers which could not be known, necessary or in general, could also be too restrictive.

A practical way for designing nonlinear control systems is provided by Fuzzy Logic It achieves non-linearity through piece wise linear approximation.

The basic building blocks of a fuzzy logic control system are fuzzy rule-based models (sets of if-else) that approximate a functional mapping.

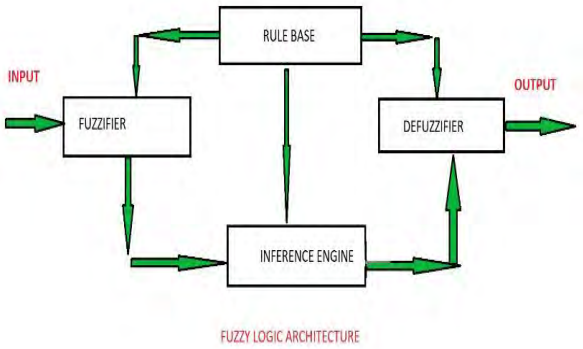


Figure 2: Components and Architecture of Fuzzy Logic

Artificial Neural Network (ANN)

ANN is a computational structure designed to mimic or copy the biological neural network. It is comprised of neurons which are connected by interconnected weights. ANN has three types of neuron namely- input nodes, hidden nodes & output nodes.

Mostly two types of ANN are used:

Feed_ Forward Neural Network

Feed_ Backward Neural Network

Feed Forward Neural Network

Feed-Forward Neural Network was the arguably simple and first neural network. In this, information flow is uni- directed i.e. forward direction. In this type of network, no loops and cycles are formed.

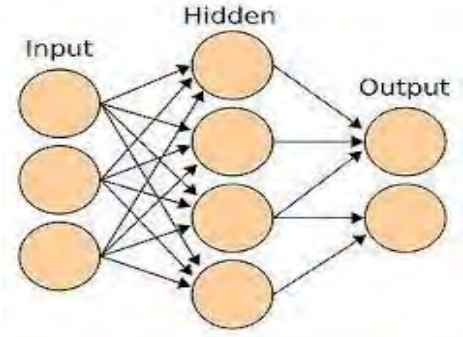


Figure 3: Feed Forward Neural Network

Feed Backward Neural Network

Feed-Backward Neural Network is also known as the recurrent neural network. Feedback connections are contained by these neural networks. In this, data flow is in both the directions. Possible of loop formation is there in this network.

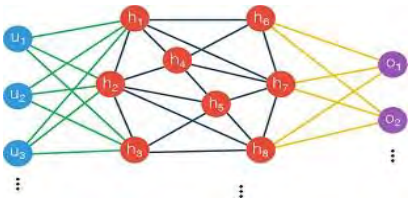


Figure 4:Feed Backward Neural Network

Problem analysis by ANNs

This fish-bone structure represents the important factors to be considered while analyzing NN model problems. An essential way of understanding the objectives sub- problems, to train the network for improving its learning by iteration.

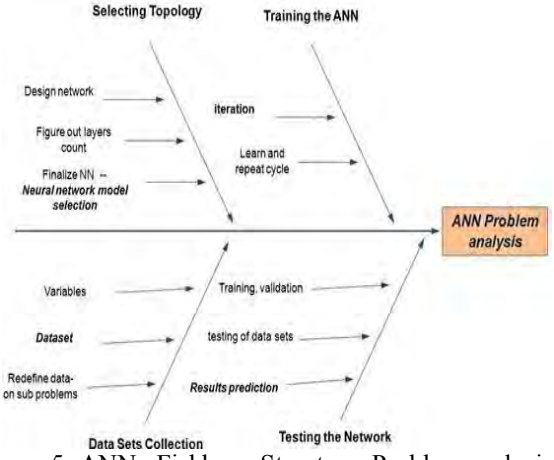


Figure 5: ANN –Fishbone Structure: Problem analysis

Genetic Algorithm (GA)

In terms of nature, evolution refers to competition among different individuals for resources in the environment or generally a natural selection. Those individuals who can survive and propagate genetic material are considered better. The diversity in the population is calculated by mutation operation.

Usually found grouped under term evolutionary computation or evolutionary algorithms, are domains of genetic algorithms (GA), evolution strategies evolutionary programming & genetic programming.

Possible better solutions are generated on the basis of common concept by simulating the evolution of individual structures through processes of selection, recombination & mutual reproduction. [11].

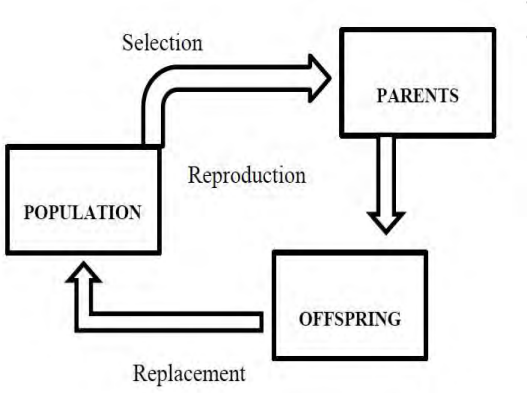


Figure 6: Flowchart for Genetic Algorithm

II. LITERATURE SURVEY

Soft computing has tremendous advancement in the engineering fields such as Aircraft, mobile, robot, cooling, and heating, power electronics etc. One of the emerging and striking examples of soft computing techniques is the "neuro-fuzzy"[1] system – a combination of neural network and fuzzy logic. These systems are becoming increasingly famous among customers, ranging from air conditioners, washing machines, photocopiers, and camcorders.

It is not easy to control the growing complexity of modern machinery by making use of traditional control systems techniques. Plants with large time delays such as nonlinear and time-variant plants cannot easily be controlled and stabilized using traditional techniques. The lack of an accurate model that describes the plant is one of the reasons for this difficulty. Soft computing is proving to be an efficient way of controlling such complex plants[2].

It has been discovered that Soft Computing could prove to be one of the efficient methods to reduce power consumption in the cloud. An N-P complete problem can be efficiently solved by using Soft Computing approach. As compared to Hard Computing, better results in this scenario are achieved using Soft Computing. Therefore, it will be preferable to use advanced techniques to follow Hybrid Soft Computing[3] approach for energy efficiency in cloud computing.

Effectiveness of Soft computing methods to detect the fault in A.C. Motor[4]. Feed Forward Back Propagation (FFBP) and Radial Basis Function (RBF), Neural Network controller are being used as a soft computing method and Three Phase Induction motor is used as an AC Motor. The healthy and unhealthy condition of the motor is tested and classified using both the neural network controllers. Both controllers give good results for the above purpose but the RBF neural network shows very encouraging results for multi-class classification problems. thereby proving the effectiveness of the controller for setting up a base in incipient fault detection of Induction Motor.

There are a number of discussions and explanations regarding the evolvement of Big Data Challenges[5], and many complexities involved in Big data mainly format and unstructured nature. These have to be organized and extracted using computational intelligence and soft computing.

III. RECENT TRENDS IN FUZZY LOGIC

Some of the recent real-world examples are stated below where Fuzzy Logic is taken into consideration and the results are highly effective.

Machine Learning and Pattern Recognition:
Several tasks are supported by machine learning techniques, such as clustering, classification, optimization, rule development and many more[6]. In these processes, data is needed to be evaluated and similar patterns are extracted and captured from raw data. Therefore, it becomes suitable for tasks like image recognition and text classification.

Medical Science:
Fuzzy set theory and fuzzy logic are a highly appropriate and relevant source for evolving knowledge-based structures in medicine for accountabilities such as the elucidation, explanation of sets of medical discoveries and outcomes, diagnosis of diseases, mixed diagnosis, the optimal selection of medical treatments, and for real-time observing of patient data [7].

Transportation Industry:
In the transport industry, the fuzzy-based system has a variety of applications ranging from traffic scheduling, braking to signaling and many more [8].

Green Energy:
Fuzzy controllers are being implemented in the system along with power electronic interfacing units to stabilize the output voltage at the load side for achieving a constant- voltage irrespective of the change in supply voltage and load in a solar photovoltaic cell-based application[9].

Data Mining and Clustering:

In order to provide useful information to the user for decision making, data mining extracts information from large databases. Dependencies between the data are discovered by fuzzy sets.
Business customer market share and telecommunication customer data mining are effectively done using Fuzzy Clustering[10] algorithms.

Psychology:
It has been analyzed that Fuzzy Logic has numerous applications in the field of Psychology[11] namely- emotion, perception, motivation, anxiety, frustration, stress, etc.

In Psychology, fuzzy logic is used in the following areas

- Fuzzy logic-based analysis of human behaviour
- Criminal investigation and prevention based on fuzzy logic reasoning and many more.

Cyber Security:
Advanced Cyber Security System (ACSS) [12] are being modelled and implemented using Fuzzy Logic, consisting of rule depository and mechanism for accessing and executing those rules. Where the depository is usually a collection of the set of Fuzzy rules

IV. RECENT TRENDS IN ARTIFICIAL NEURAL NETWORK

E-Learning through ANNs:
ANN supports application scope to implement and utilize the neural model for the personalization level. Smart e-Learning applications [13] must require neural system design features for learning behaviour analysis to utilize for betterments of academic teaching and learning processes.

Forecasting:
Everyday business decisions (e.g. sales, the financial allocation between products, capacity utilization), require forecasting extensively in economic and monetary policy, in the finance and the stock market. Forecasting problems are complex, more often. Predicting stock prices is a complex problem with a lot of underlying unknown and unseen factors. There are limitations in traditional forecasting systems in terms of taking into account these complex, non- linear relationships. A robust alternative can be provided if ANNs are applied in the right way, giving it ability to model and extract unseen features and relationships. Also, ANN doesn't impose any restriction on input and residual distributions unlike the traditional models.

Signature Verification Application:
The primary approach is extracting the feature or the geometrical feature set that represent the signature.Using these feature sets,train the neural networks by making use of an efficient neural network algorithm. Classify the signature by this trained neural network as being genuine or forged under the verification stage.

Human Face Recognition:
All the input images need to be preprocessed, at first. Then, the dimensions of that image needed to be reduced. And, at last, classification is done using a neural network training algorithm. Following are the neural networks used for training purposes with preprocessed image −

Fully-connected multilayer feed-forward neural network is trained with the help of the back- propagation algorithm.

Principal Component Analysis (PCA) technique is used, for dimensionality reduction.

V. RECENT TRENDS IN GENETIC ALGORITHM

Application domains in Genetic Algorithm
New applications where GAs prove to be well-suited tools are continuously found. Bioinformatics, image analysis, signal processing, telecommunication, hardware optimization, and artificial music and art represent recent application domains [14].

Genetic Programming
Since beginning of the early 90s, much progress has been witnessed with respect to genetic programming. This class belongs to genetic algorithms, uses programs in the form of tree-structures as required solutions. It has the ability to automatically produce results that are competitive to be patented or human inventions that can be patented[13].Such successful applications are analyzed and reported for the design of electronic circuits for analog filters and controllers, and for the development of quantum algorithms, and certain problems in bioinformatics.

Extended hybrid architectures:

For solving discrete optimization problems, classical simple EAs are found to be typically less efficient compared to other state-of-the-art algorithms that are specifically tailored to the problems. As a result, EAs are often combined with problem-specific algorithms. The hybridised form of EAs and other algorithms such as local search, simulated annealing, guided local search, variable neighbourhood search, and variable depth search are widely used in recent time, and for several hard-combinatorics optimization problems, such techniques are today's topmost algorithms for computing superior solutions to large problem instances within limited time. Such combinations are also known as memetic algorithms. Typically, the EA's primary responsibility is to need care of diversification and coverage of all major areas of the search space, while the embedded other metaheuristic is particularly responsiblefor the intensification of the search, thus, the fine-tuning of the foremost promising candidate solutions.

Estimation of distribution algorithms (EDAs):
In this relatively new class of EAs, a probabilistic model is used to represent desired solutions. A new generation of candidate solution is calculated by taking samples of the probabilistic model. These solutions are evaluated, and a subset is selected which later is used to update and refine the probabilistic model. *Univariate marginal distribution algorithm* (UMDA) is the simplest form of an EDA. In it, the variables which represent a solution are regarded as independent and the probabilistic model is a vector of associated probabilities. More sophisticated EDAs also consider Dependencies among the variables is also considered by more sophisticated EDAs, Bayesian networks as probabilistic models is one of the examples.

Neural Networks: Neural networks, in particular recurrent neural networks are trained using GAs.

Sl.no.	Field of applications	Soft computing components
1	Aircraft and air traffic	NN, FL, EC
2	Communication networks	FL, NN, EC
3	Control and Monitoring	EC, FL, NN
4	Cooling and Heating	FL, NN, EC
5	Data communications	FL, NN
6	Data Security	ANN, FL
7	Induction Motor Drives	FL, NN
8	Inverters and Converters	FL, NN
9	Manufacturing Technologies	FL, NN
10	Mobile Robots	FL, NN
11	Multi-Agent Robots	EC, FL
12	Network Optimization	GA
13	Power Control	EC
14	Radio Planning	ANN
15	Resource Allocation	ANN
16	Satellite Imaging	ANN, FL, EA
17	Scheduling	ANN
18	Spacecraft	NN, FL,
19	Steel Process Industry	FL, NN
20	Switched Reluctance Motor Drives	FL

Table 1: Single and Multiple usage of SC technique in various applications[15].

VI. CONCLUSION

The evolvement of soft computing techniques is remarkable in every field of physics, material sciences, computer, chemistry, statistics, etc. Outline of various soft computing techniques and their different applications and techniques have been represented in this paper. The robustness, cost-effectiveness, simplicity are some of unique and remarkable features of soft computing. Fuzzy logic is very suitable for tracking imprecision and uncertainty and has wide variety of applications. Whereas ANN offers good prediction capabilities thanks to its generalization and learning features. On the other hand, Evolutionary computing techniques are very much applicable in optimization problems and for estimation of different distribution algorithms. As the development of soft computing flourish day by day, the application areas will also be felt increasingly in the coming years. Day by day Soft Computing based products are increasing. Majority of such products uses any of the soft computing technique inside the sub-systems which aren't known to end-user. The gist is that soft computing techniques will become common to various applications and can deal with imprecise problems.

VII. References

[1] Kurhe A.B., Satonkar S.S., Khanale P.B., Shinde Ashok, "Soft Computing and its Applications", BIOINFO Soft Computing, Volume 1, Issue 1, 2011.

[2] Mrs.S.Geethamani,Mrs.V.Uthra,Ms.V.Ramya,Mrs.B.Pavithr, "A STUDY ON SOFT COMPUTING AND APPLICATIONS", International Journal of Scientific Research and Review, Volume 7, Issue 3, 2019.

[3] Kumari Parul Jasuja, ER.Khushdeep Kaur, "Hybrid Soft Computing Approach for Energy Efficiency in Cloud Computing", International Conference on Communication and Electronics Systems (ICCES), 2016.

[4] Pratap Sekhar Puhan, Sudersan Behera, "Application of Soft Computing Methods to Detect Fault in A.C Motor", International Conference on Advances in Computing, Communication and Control, 2017.

[5] Ahmed Ghulamrasul Al Balushi, Dr. Herald Noronha, Dr. Prakash Kumar U, "Evolvement of Big Data Challenges in Computational Intelligence and Soft Computing", 2019.

[6] Aruna Bajpai, Virendra S. Kushwaha, "Importance of Fuzzy Logic and Application Areas in Engineering Research",Volume7, Issue 6, 2019.

[7] Rahul Isola, Rebeck Carvalho, Amiya Kumar Tripathy, "Knowledge Discovery in Medical Systems Using Differential Diagnosis, LAMSTAR, and k-NN", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, Volume 16, Issue 6,2012.

[8] Amrita S, G. Sahoo, U.C.Sahoo, "Application of Fuzzy Logic in Transport Planning", International Journal on Soft Computing, Volume 3, Issue 2, 2012

[9] Prakhar Sharma, Pankaj Kumar, Nitai Pal, "Design and Implementation of Closed Loop Controlled Boost Converter for Solar Power Installation System using Fuzzy Logic Controller", 4th International Conference on Recent Advances in Information Technology, 2018.

[10] Menal Dahiya, "Applications of Soft Computing in Various Areas", International Journal of Engineering Sciences and Research Technology, 2017.

[11] Shilpa Srivastava, Millie Pant, Namrata Agarwal, "A Review on Role of Fuzzy Logic in Psychology", Proceedings of Fifth International Conference on Soft Computing for Problem Solving,2016.

[12] Javed Alam, Dr. M. K. Pandey, "Advance Cyber Security System using fuzzy logic", ACME: Journal of Management &IT, Volume 10, Issue 1,2014.

[13] M.R.M.Veera Manickam, M. Mohanapriya, S. A. Kale, Mithapalli Uday,Prashant Kulkarni, Yuvraj Khandagale, Suraj P Patil, "RESEARCH STUDY ON APPLICATIONS OFARTIFICIAL NEURAL NETWORKS AND E-LEARNING PERSONALIZATION", International Journal of Civil Engineering and Technology (IJCIET)Volume 8, Issue 8, 2017.

[14] Gunther Raidl, "Evolutionary Computation: An Overview and Recent Trends", Austrian Science Fund (FWF) under grant P16263-N04.

[15] Santosh Kumar Das , Abhishek Kumar , Bappaditya Das, A.P.Burnwal, "ON SOFT COMPUTING TECHNIQUES IN VARIOUS AREA", Rupak Bhattacharyya et al. (Eds) : ACER, pp. 59–68, 2013.

Design and Development Automatic Gardening System using web API & Fuzzy logic

Mrs. Rashmi Thakur, Kunal Singh, Vignesh Singh, Pradyuman Sharma
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
rashmi.thakur@thakureducation.org, kunal4104@gmail.com, vigdrusingh@gmail.com, pradys.sharma10@gmail.com

Abstract— **the amount of water required by a plant depends upon number of factors, which include plant's species, soil type, weather and various other micro parameters. In general, best way to water most plants is by applying enough to moisten the plant's entire root system, and then letting the soil dry out slightly before watering again. A soil moisture sensor can be used to check the soil near plant roots and water them accordingly. But, plants can receive water naturally through rain and as we know, too much water can be as harmful as too little water to the plants. In this project we are planning to build an automatic watering system which not only measure soil moisture but also check the weather depending on which it makes decision to water the plant. To check the weather we will be using a web based API which will communicate with our Adriano microcontroller. If, in near future there is chance of rainfall the system will refrain from watering the plants regardless of soil moisture status. This project will help reduce the human effort required and also decrease water consumption, while maintaining the health of the plants.**

I. INTRODUCTION

Nowadays there is scarcity of water all over the world, especially in India due to low groundwater table and irregular rainfall most of the water gets wasted in drainage. Thus there is a need of more efficient and cheap alternative to traditional form of watering the plants.

Automated plant watering system estimate and measure the existing plant and then supplies desired amount of water needed by that plant. It is minimizing the excess water use as well as keeping plants healthy. This system

Saves time and is accurate. At places like societies and Building gardens we can implement an automatic watering system which is easy and accurate for the growth of plants. This will minimize the human intervention and will also save time.

To demonstrate building and enhancing of existing automatic plant watering systems, we exploit available technologies we have such as a free web hosting and

Free cloud to develop such a system. We believe that our

Proposed system can be applied to different types of plants.

II. BACKGROUND

Fuzzy logic Natural language (like most other activities in life and indeed the universe) is not easily translated

Into the absolute terms of 0 and 1. (Whether everything is ultimately describable in binary terms is a philosophical

Question worth pursuing, but in practice much data we might want to feed a computer is in some state in

Between and so, frequently, are the results of computing.)

It may help to see fuzzy logic as the way reasoning really works and binary or Boolean logic is simply a special case of it.

Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human- to-computer interaction

Embedded System is combination of computer hardware and software which is designed for a specific function. Industrial machines, agricultural and process industry devices, automobiles, medical equipment, cameras, household appliances, airplanes, vending machines and toys as well as mobile devices are all possible locations for an embedded system. Embedded system technology reduces the cost and size of electronic circuit for project developments.

III. PROPOSED IDEA

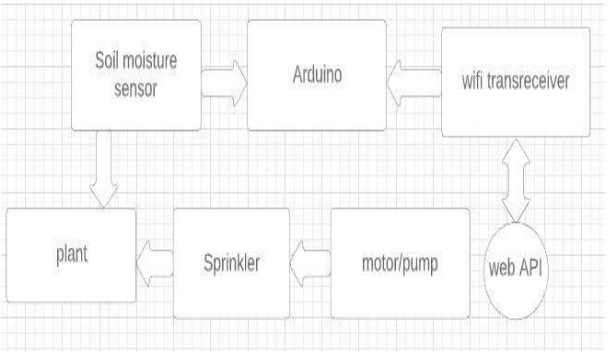


Figure 1: Block diagram

Our project is based on three functional model. First function is to measure the moisture content in soil. Second is to keep track of weather conditions by the use of real time data from open source Weather API. Lastly analyses both the data, compare to pre-set values for the given plant and decide whether to start the motor for water supply. The complete system is controlled by Adriano board which is programmed using Adriano IDE software.

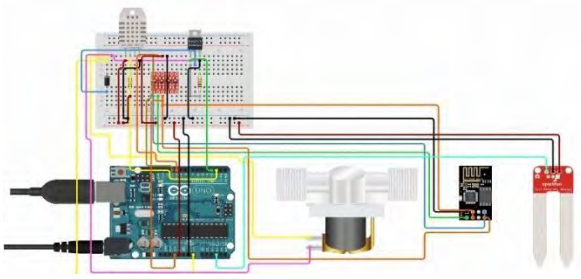


Figure 2: Pictorial Representation of System

A. Detecting soil moisture content:
The measurement of water content in the soil is achieved by usage of soil moisture sensor which sends an alert to the microcontroller when the soil moisture drops below
Threshold level.

B. Gathering Weather data
To track the weather conditions of near future we use open weather map API which is an open source weather
API that provides data in JSON format to the microcontroller in every ten minutes using Wi-Fi Transceiver. This data includes mean temperature, probability of shower, air humidity and wind speed.

Automatic watering and recording data
Based on the analysis and comparison of collected data with certain preset requirements and fuzzy logic
Algorithm the system determines the appropriate conditions to start the motor and water the plants

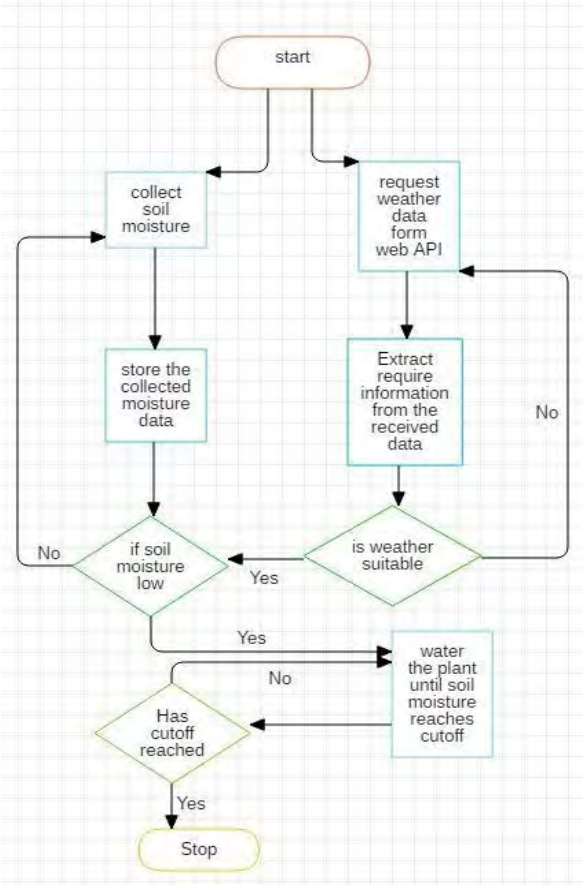
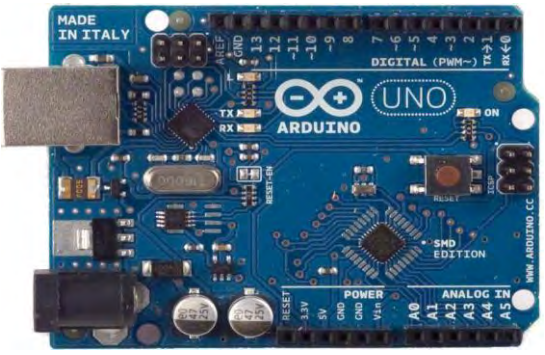


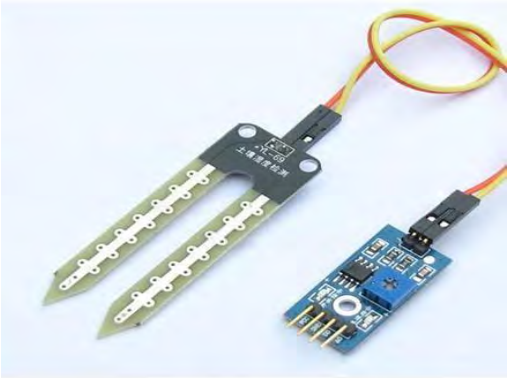
Figure 3: Flowchart of Automatic Watering System

IV. COMPONENTS USED FOR SYSTEM

1. Adriano Uno: This microprocessor is used to gather collected data from adjoining modules and takes decision to start watering.



2. Soil Moisture Sensor: Soil moisture sensor will be used to measure moisture content in the soil.



3. ESP8266-01 (Wi-Fi Transreciever): Tran's receiver will be used to collect data from openweathermap.org API to get real-time weather reports.

4. Relay: The main operation of this device is to make or break contact with the help of a signal without any human involvement in order To switch it ON or OFF. We will be using this to Control the motor to pump water.

5. Water pump: DC powered pumps use direct current from motor, battery, or solar power to move fluid in a variety of ways.

6. SD card module: This module will be used to collect data in the SD card for future analysis.

V.RESULT AND ANALYSIS

During the process of testing we have achieved positive results from this project. It was found that our plants were able to grow without dehydration whilst using limited water. Although, further optimization can be Done to our fuzzy rule set. The data stored by SD module can be used to further analyses the pattern using data analysis algorithms like linear regression, K-mean clustering etc. We can also add a mini camera module available with esp32 Wi-Fi module to store growth process in the SD card which can be converted to a time-lapse to further your understanding of the gardening.

V.CONCLUSION

Design and develop a smart system that can sense the content of water in the soil. And by analyzing the results irrigation becomes easy, accurate and practical. The Volumes, volumetric water contents play an important role in producing the output.

The primary application of this project is to provide plants with adequate water without wasteful use of water and manual labor. This project is making the use of fuzzy inference system to decide the plant

Watering process. This project can further be extended to larger scale combining with the principles of rain water harvesting; it could be instrumental in areas with severe shortage of rainfall.

The principle can be extended to create fully automated gardens and farmlands. Combined with the principle of rain water harvesting, it could lead to huge water savings if applied in the right manner. In agricultural lands with severe shortage of rainfall, this model can be successfully applied to achieve great results with most types of soil.

This Smart irrigation proves to be the system automates for irrigation system and regulates water for irrigation is done without manual Using this system, solenoid valves and relay board can be controlled remotely which opens the opportunities to control the water flow as well as the electrical flow. Irrigation system is automated with depends on sensor Report the pump is operated by the weather condition by soil, rain and temperature conditions the water pump will work and by wireless trans receiver the data is communicate and the sensor readings are uploaded in

to cloud network by Wi-Fi___33 technology

.

Deep Learning Using Boltzmann Machines & Belief Network

Tarunima Mukherjee , Ganguly,Prof.Veena Kulkarni
Thakur College of Engineering & Technology
Kandivli East – Mumbai, India

Abstract: **- In recent times, the field of deep learning has achieved great importance and success in fields like natural language processing and computer vision. Deep Learning with its strong ability to learn by making use of datasets for feature selection is gaining more importance than traditional machine learning. Deep learning is becoming popular amongst researchers due to its practicability. In this paper we discuss about some of the advanced neural networks with its applications. We also, discuss about the limitation and advantages of deep learning with Deep Boltzmann Machine (DBM) & Deep Belief Networks (DBN) being the area of study.**

Keywords – Deep Learning, DBM, DBN, Neural Network, Dataset

I. INTRODUCTION

Deep learning is a machine learning technique which helps the computers to learn by example. Images, text, or sound helps the computer model to perform classification tasks directly. High degrees of accuracy, sometimes exceeding human-level performance is achieved by Deep Learning. The deep learning models are trained by using a large set of labeled data and neural network architectures that contain many layers.

Discovering high-level representations from high dimensional data is the basic idea of resolving multiple AI related tasks, which includes object recognition, speech perception and language understanding. To create such systems, we require deep architectures which includes many layers of nonlinear processing.

The Greedy algorithm suffers from various disadvantages such as
1) It is based on a very approximate inference procedure, limited to a single bottom-up pass. 2) The model can fail to effectively account for uncertainty while interpreting ambiguous sensory inputs. 3) The existing greedy procedure is suboptimal as it learns one layer of features at one time and never re-adjusts its lower- level parameters. 4) While global fine-tuning has been done by using the contrastive wake-sleep algorithm of Hinton et al. (2006) still it turns out to be very slow and inefficient.

Deep Boltzmann Machine (DBM) is a new kind of learning algorithm that was introduced by Hilton & Salakhutdinov for hierarchical probabilistic model. DBM is a type of Markov random field, with undirected connections between layers. The concept of Deep Boltzmann Machines are exciting for several reasons. Firstly, it retains much of the desired data as found in Deep Belief Networks. Secondly it also discovers several layers of increasingly complex representations of the input data and comes with an efficient layer-by-layer pretraining procedure. Thirdly it can also be trained on unlabeled data and can be tweaked for a specific task using the labeled data. For DBN's and deep convolutional neural networks, the approximate inference procedure for DBM's incorporates a top-down feedback in addition to the usual bottom-up pass, thereby enabling Deep Boltzmann Machines to better incorporate uncertainty about ambiguous inputs. Also, the parameters of all layers can be optimized jointly by following the approximate gradient of a variation lower- bound on the likelihood function. Greatly facilitates learning thereby providing better generative models

II. MODELS OF DEEP ARCHITECTURE

Deep Architecture learning algorithms are the one that learn to represent functions with many levels of composition. It is made up of many layers of Non-Linear Operations like neural nets with multiple hidden layers. The building blocks of deep architectures such as DBN and DBM are the RBMs.
Now we will look at some of the deep learning models: -

A) Deep Boltzmann Machine (DBM)
DBM is a network of stochastic binary units coupled symmetrically. It contains a set of visible units $v \in \{0, 1\}$, and a sequence of layers of hidden units $h1 \in \{0, 1\}F1$, $h2 \in \{0, 1\}F2$,..., $hL \in \{0, 1\}FL$. There are connections only between hidden units in adjacent layers, as well as between the visible units and the hidden units in the first hidden layer [1]. This is represented in Figure 1[2].
The various advantages include the following points: -

Advantages:
Firstly, it retains much of the requisites found in Deep Belief Networks. Secondly, it discovers several layers of increasingly complex representations of the input, and it comes with an efficient layer-by-layer pre-training procedure. Also, it can be trained on unlabeled data and can be modified for a specific task using the labeled data.

Also, the approximate inference procedure for DBM's incorporates a feedback which is top down in nature in addition to the bottom-up pass, which allows Deep Boltzmann Machines to include uncertainty of ambiguous inputs. Additionally, all the parameters of all layers can be jointly optimized by following the approximate gradient of a variation lower-bound on the likelihood function which in turn enables better learning generative models.

2) Disadvantages: -

One of the disadvantages is described as below: -

The approximate inference is based on the mean-field approach, which is considerably slow when compared to a single bottom-up pass of Deep Belief Networks. Due to this disadvantage the joint optimization of DBM parameters becomes unfeasible for large datasets. Secondly, this also reduces the appeal for using DBM's in extracting useful feature representations as the expensive mean-field inference has to perform for every new test input.

3) Characteristics of DBM

- The various features of Deep Boltzmann Machine are as below:-These are unsupervised, generative & probabilistic model with completely undirected connections between various layers
- Similar to RBM, no intralayer connection exists in DBM.
- In DBM connections exists only between units of the neighbouring layers
- This contains visible & multiple layers of hidden units.
- DBM is made up of binary units which are symmetrically connected and are stochastic
- DBM can be seen as bipartite graph with odd layers as one set of vertices and even layers as another set

.

- In DBM the units within the layers are dependent on neighbouring layers but are independent of each other.
- Efficient learning is ensured by pre training of all the layers one at a time. However, it is to be noted that Greedy layer wise pre training is different than DBM pretraining.
- By using back propagation, the DBM layers can be fine-tuned.

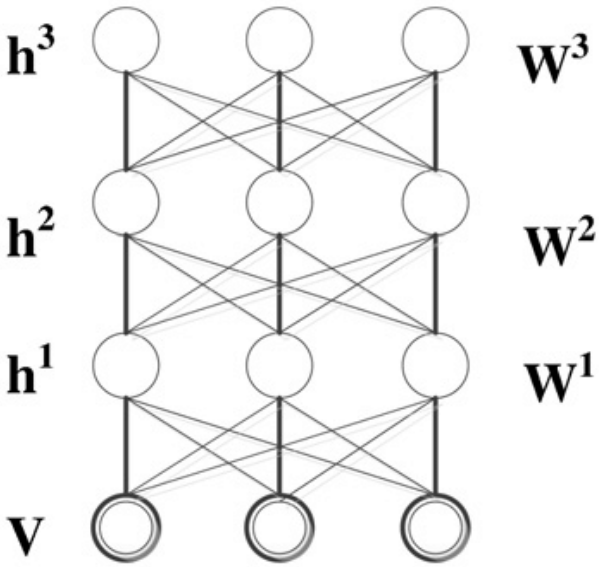


Fig. 1 Graphical Representation of DBM

B. Deep Belief Network (DBN)

The second model which will be discussed is the Deep Belief

Net (DBN) by Hinton [3].

DBN is obtained by training and stacking several layers of Restricted Boltzmann Machines (RBM) in a greedy manner. In this the stack of RBMs is trained at first and then it is used to initialize classification of multi-layer neural network.[4].

DBN is a multi-layer generative model with layer variables as h0 (the input or visible layer), h1, h2, h3 etc. The top two layers have a joint distribution which is an RBM, where P(hk|hk−1) is parametrized in the same way as it is done for RBM. Therefore, we sat that a 2-layer DBN is an RBM, and a stack of RBMs share parametrization with a corresponding DBN as in Fig 2. [5]

Thus, we see that Deep belief nets are probabilistic generative models and is made up of multiple layers of latent variables having binary values and are termed as feature detectors or hidden units. Also, we see that the undirected and symmetric connections of top two layers forms an associative memory. To initialize each layer of a DBN as an RBM, the contrastive divergence update direction is used as below: -

- The first layer of the DBN is trained as an RBM P1 with visible layer v1 and hidden layer h1.
- The second layer RBM P2 is trained such that the models with its visible layer v1 is sampled from the training data set and the sample h2 from P1(h1|v1).

It has been shown in [6] that this maximizes a lower bound on the log-likelihood of the DBN. With the newly added top layer trained as an RBM to model the samples produced by chaining the posteriors P(hk|hk−1) of the lower layers

(starting from h0 from the training data set) the number of layers are increased greedily, [6].

1) Greedy layer wise training

Greedy layer wise training is accomplished by first training the lower layer with an unsupervised learning algorithm, thereby giving rise to an initial set of parameter values for the first layer of a neural network. Then the output of the first layer is used as an input for another layer, and similarly initialize that layer with an unsupervised learning algorithm. After having thus initialized a number of layers, the whole neural network is fine-tuned with respect to a supervised training criterion. [4].

2) Advantages

Firstly, one of the advantages of the greedy layer-by layer learning algorithm is that it can quickly find a find a good set of model parameter, even if the models contains huge number of parameters. Secondly, the learning algorithm efficiently use large sets of unlabeled data. Thirdly, the model can be pre-trained in a completely unsupervised fashion. Lastly, there is an efficient way of performing approximate inference to compute the values of the latent variables in the deepest layer with some given input [7].

3) Disadvantages

One of the main disadvantages of the greedy algorithm is that the approximate inference procedure is limited to a single bottom-up pass. The model fails to adequately account for uncertainty when interpreting ambiguous sensory inputs because of ignoring top- down influences on the inference process. Also, as the greedy procedure learns one layer of features at a time and never re-adjusts its lower-level parameters thereby resulting in the making the greedy procedure as sub optimal. Despite global fine- tuning using the contrastive wake-sleep algorithm has been used by Hinton, still the greedy algorithm it is very slow and inefficient [7].

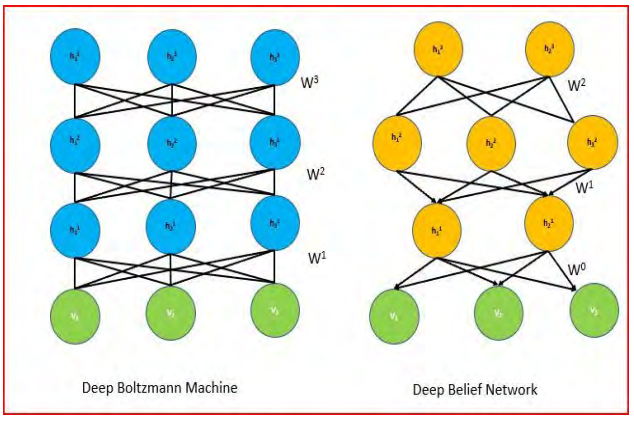


Fig. 2 Graphical Representation of DBN

C Deep Belief networks (DBN) vs Deep Boltzmann Machine (DBM)

- For Deep Belief Network (DBN) the top two layers are having undirected connections while directed connections exists for the lower layers. On the other hand /.the Deep Boltzmann Machine (DBM) are having entirely undirected connections.
- A top-down feedback along with the bottom-up pass is used for approximate inference procedure for DBM, thereby allowing Deep Boltzmann Machines to have better control over uncertainty about ambiguous inputs.
- The approximate inference based on mean field approach as used in DBN is slower as compared to a single bottom-up pass as used in Deep Belief Networks. This is because for every new input, the Mean field inference needs to be performed again and again thereby reducing its efficiency. The differences are shown in Fig 3 [8]

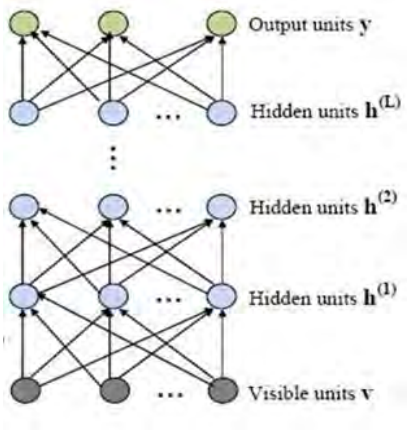


Fig. 3 Graphical Difference between DBM & DBN

The table below lists down the major differences between DBM & DBN Table I. Comparison of DBN with DBM

S.No	Deep Boltzman Machine	Deep Belief Network
1	DBM are undirected	Belief Networks are directed
2	Each layers forms an RBM because the connection between all layers are undirected	The first 2 layers forms a RBM then subsequent layers form a directed
3	These are Markoff Random Fields	These are sigmoid belief networks
4	Computing P (v | h) is infeasible	Computing P (v | h) is easy
5	Intractable Function makes P infeasible	No Intractable Partition Function

6	DBM gives lesser error rate as compared to DBN on various	DBN has higher error rate as compared to DBM.

D. Applications of the Model

In this section we will be discussing about various applications of Deep Boltzmann Machine and Deep Belief Network.

Deep Boltzmann Machine:-

1) Shape Completion: - It is an important task in the field of image processing. DBM with its strong ability to deal with distribution of shapes, it is easy to achieve the result by sampling from the model. Here hidden activation of the DBM is used and incorporated with the convolutional shape features to fit a regression model and then the output of the regression model is compared with the incomplete shape feature in order to set a mask for sampling from the DBM.

2) Deep Feature Selection - The structure of a Deep Boltzmann Machine enables it to learn very complex relationships between features and facilitates advanced performance in learning of high- level representation of features. Deep Feature Selection (Deep- FS), is used to remove irrelevant features from large datasets in order to reduce the number of inputs which are modelled during the learning process. Deep-FS algorithm uses a Deep Boltzmann Machine, to remove features at the initiation of the learning process.

3) Object Recognition:- Object Recognition means to detect and classify objects in an image or video into generic categories DBM uses binary hidden and visible units, which is applicable to quasi- binary images (e.g., handwritten digits). The extension of DBMs are more suitable to the continuous data. The characterization of object recognition using DBM has been a recent focus in the computer vision community.

4) DBMs are extensively used in solving combinatorial optimization problems.

5) DBMs are also used for detecting Parkinson Disease using signal decomposition.

Deep Belief Network:-

1) Generating and recognizing images-In this the input will be a picture and the output will be a category. It has a wide range of application varying from simple photo organization to critical functions like medical diagnoses. For example, smart microspores which can do image recognition could be used to classify pathogens.

2)Video sequences -It works similarly to vision, in which it finds Meaning in the video data. For example, it can be used to detect an object or a gesture of a person. It can also be used in many different fields such as home automation, security and healthcare.

3) Motion-capture data - Motion capture data involves tracking the movement of objects or people by using deep belief networks. Motion capture can easily lose track and depends on what an object or person appearance velocity and distance. It is used in video game development and in filmmaking

4) Natural language understanding problem- It is done by using an efficient learning algorithm called contrastive divergence (CD). CD enables DBNs to learn a multi-layer generative model from unlabeled data and the features discovered are then used to initialize a feed-forward neural network which is fine-tuned with backpropagation

5) Precision Mechanism Quality Inspection- Precision mechanism is widely used for various industry applications. Quality inspection for precision mechanism is important for manufacturers to ensure that the product dispatched is of good quality.This is achieved by a detection method, based on a Deep Belief Network (DBN) auto-encoder known as Tilear.It is structured in two parts: training and decision-making. Tilear is trained by using the signals only from good samples, which in turn enables it to reconstruct signals of good sample only. If the signal from the sample part matches the Tilear reconstructed signal, then the part is of expected quality and facilitates decision making.

6) Image compression and extraction -This application powers the capacity of deep belief network for classification of raw data from their features and increases their ability to handle large number of parameters.

III. CONCLUSION

When it comes to representation learning techniques, we find that Deep learning is one of the most powerful tool. In this paper, it has been represented as to how the ability of one deep learning architecture i.e. the deep belief network varies from the Deep Boltzmann machines.

Deep Belief Networks (DBNs) and Deep Boltzmann Machines (DBMs) look very similar diagrammatically, however they are actually very different qualitatively. Also, both the techniques are probabilistic graphical models composed of stacked layers of RBMs with the difference being in the manner these layers are connected with each other.

For both the models the learning tasks are enabled by using greedy layer-by-layer training procedure. The greedy layer-by-layer learning algorithm very quickly finds a good set of model parameters, even if the models contains many

parameters and layers of non-linearities. Also, the model can be pre-trained in a completely unsupervised manner. A grave disadvantage of this greedy algorithm is that it is based on an approximate inference procedure which is restricted to a single bottom-up pass. Compared to DBNs, DBMs are found to be more effective. Unlike existing models with deep architectures, including DBNs and deep convolutional neural networks, the approximate inference procedure for DBMs incorporates a top-down feedback along with bottom-up pass, enabling DBMs to include uncertainty about ambiguous inputs, also the easy implementation enables its usage in various data set as compared to the intricate models of deep convolutional neural networks Third, and perhaps more importantly, parameters of all layers can be improved jointly by

.

following the approximate gradient of a variational lower-bound on the likelihood function. This greatly facilitates learning better generative models. However, a critical weakness of DBMs is that approximate inference, which is based on the mean-field approach, is considerably (between 25 and 50 times) slow as compared to a single bottom-up pass as in Deep Belief Networks. Due to this for large datasets the joint optimization of DBM parameters becomes impractical

IV.References

[1] Geoffrey Hinton, Ruslan Salakhutdinov, *"Deep Boltzmann machine",* In 12th International Conference on Artificial Intelligence and Statistics (AISTATS), Volume 5 of MLR,Florida, USA, 2009.

[2] www.researchgate.net/figure/Deep-Boltzmann-Machine-with- three-hide-layers_fig2_315303699

[3] Yoshua Bengio,*"Learning deep architecture for AI",Foundation and trends in Machine Learning,Volume 2* Issue1,pages 1-127, January 2009

[4] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, *"Why does unsupervised pre-training help deep learning? ", Journal of Machine Learning Research, JMLR,*

[5]www.slideshare.net/HasanHTopcu/deep-belief-networks-58155447

[6] Bengio, Y., & LeCun, Y., *"Scaling learning algorithms towards AI",* In Bottou, L., Chapelle, O., DeCoste, D., & Weston, J. (Eds.), Large Scale Kernel Machines. MIT Press, 2007.

[7] Ruslan Salakhutdinov, Hugo Larochelle , *"Efficient Learning of Deep Boltzmann Machines ",Neural computation, ,Volume 24*,No.8, Pages 1967-2006, August 2012

[8]medium.com/datadriveninvestor/deep-learning-deep-boltzmannmachine-dbm-e3253bb95d0f

Study of Predictive Analysis in Fin Tech

Pooja sharma, Dr. Sheetal Rathi
Dept. of Computer Engineering
Thakur College of Engineering Technology
pooja12sharma.1997@gmail.com, sheetal.rathi@thakureducation.org

Abstract-***One of the leading trend among investors is FinTech related venture. For discussion, people will always take an issue on financial and banking on the phones and many customers want to resolve the matter without directly going to the banks. Thousands of customer calls done for queries will make use of call center receive each day which will increase the growth of online banking and all the things now switching towards online system which will reduce cost, human error etc. and it will change the way of consumer markets and our societies. Based on predictive analytics businesses are identifying the patterns and trends to increase the efficiency of Data Mining (DM) which we proposed by using an Artificial Neural Network (ANN)approach for predicting the success of telemarketing calls in selling long term deposits of bank which will perform other conventional classifier and confirms its valuable for campaign on telemarketing. Its goal is to predict if the client will subscribe a term deposit.***

Keyword- Bank Telemarketing, Decision making, Prediction, Input parameters, machine learning techniques.

I. INTRODUCTION

Telemarketing is a interactive and personalized method of marketing the product through various medium like phone or a subsequent face to face or web conferencing appointments scheduled during the call. In banking sector telemarketing has gained an importance as an effective tool for promotion. [1]

Telemarketing having a process for making use of sophisticated communication network for promoting the banks. This includes promoting through various mediums like television, telephone, radio and nowadays very largely through cell phones. Now a today Banks have started using 'SMS' and many other services for their customers supported by cell phones to provide benefits and thus have tried to increase their sales. In today's competitive and modern scenario it very important that banks makes use of telemarketing techniques very efficiently to have desirable results [2]. Marketing selling is to promote through various mediums like TV, online, radio, media etc. done a campaigns constitute to enhance business.

Customers by meeting and contact them for a specific goal. Marketing put into operation through a contact Center is called telemarketing due to the remoteness characteristic. To predict a non-linear nature of stock there are several methods are applied. For prediction a very powerful tool is ANNs and one of the type of ANNs is Multi-Layer Perceptron (MLP). It will used to explain and predict one or more observable and effectively measured phenomena. For all variables the analysis is done on database using a supervised learning. From hidden information we explore a new information, in this case there is a target variable to be predicted. There is available data for research and was collected from a Portuguese retail bank.

II. LITERATURE REVIEW

For stock market prediction selection of input features is important task. Technical indicator were used as an input for machine learning techniques. Here further we see some of the techniques used by researchers used some input features for prediction.

Yan LeCunn is said to be one of the fathers of deep learning, mostly due to his work with a pioneering 7-Level CNN known as LeNet-5 in the year 1998 [3]. His work was applied by various banks in order to recognise hand written digits on cheques. Whilst still using the CNN Since the introduction of ANNs, other techniques such as SVM have been given more importance and have achieved a better results in a short time period.[4] In bank telemarketing Using CNNs whereby its aim is to predict whether a customer will take up a particular marketing campaign based on a number of numeric and nominal features per customer. Amongst all accuracy an impressive 76.70%, is the highest from among all 7 classifiers [5].

Y. Kawasaki & M. Ueki have examined predictive modelling with several sparse regression methods for bank telemarketing

Success they have concluded that an effective predictive model may help in reducing costs for marketing in companies. [6]

H. A.Elsalamony has estimated performance on three statistical measures: specificity, classification accuracy, sensitivity. He has used a Multilayer perceptron neural network (MLPNN) and Ross Quinlan new decision tree model (C5.0) using a database of bank to increase the campaign effectiveness by identifying the main characteristics that affect a success [7].

III. THE PROPOSED APPROACH

The direct marketing campaign of "Portuguese bank dataset" is used in this paper. [8]

A. Dataset

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribe. This research focus on targeting through telemarketing phone calls to sell long-term deposits.

Within a campaign, the human agents execute phone calls to a list of clients to sell the deposit (outbound) or, if meanwhile the client calls the contact- center for any other reason, he is asked to subscribe the deposit (inbound).[9] Thus, the result is a binary unsuccessful or successful contact. For evaluation purposes, a time ordered split was initially performed, where the records were divided into training (four years) and test data (one year). The training data is used for feature and model selection and includes all contacts executed up to June 2012, in a total of 51,651 examples. The test data is used for measuring the prediction capabilities of the selected data-driven model, including the most recent 1293 contacts, from July 2012 to June 2013.Each record included the output target, the contact outcome ({"failure", "success"}), and candidate input features. These include telemarketing attributes (e.g., call direction), product details (e.g., interest rate offered) and client information (e.g., age). These records were enriched with social and economic influence features (e.g. Unemployment variation rate), by gathering external data from the central bank of the Portuguese. Republic statistical web site. The merging of the two data sources led to a large set of potentially useful features [10].

Table 1: client attribute's information

Sr.No	Attribute's	Description
1	Age	numeric
2	Job	type of job (categorical: 'admin.', 'blue collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self employed','services','student','technician','unemployed','unknown')
3	Martial	marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)
4	Education	(categorical:'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional course','university degree', 'unknown')
5	Default	has credit in default? (categorical: 'no', 'yes', 'unknown')
6	Balance	categorical: 'no','yes', 'unknown'
7	Housing	has housing loan? (categorical: 'no', 'yes', 'unknown')
8	Loan	has personal loan? (categorical: 'no','yes', 'unknown')
9	Contact	contact communication type (categorical: 'cellular', 'telephone')
10	Day	last contact day of the week (categorical: 'mon', 'tue', 'wed', 'thu', 'fri')
11	Month	last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
12	Duration	last contact duration, in seconds (numeric).
13	Campaign	number of contacts performed during this campaign and for this client (numeric, includes last contact)
14	Pdays	number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
15	Previous	number of contacts performed before this campaign and for this client (numeric)
16	Poutcome	outcome of the previous marketing campaign (categorical: 'failure', 'non-existent','success')
17	Y/N	has the client subscribed a term deposit? (binary: 'yes', 'no')

B. Dataset Balancing, Training, and Test Split

We will be utilizing data from 2003 till 2013 (11 years) as our training set and 2014 till 2016 (3 years) as our test set. This ensures a substantial amount of data for training, whereby 11 years of stock movements should cover a wide range of long and short term trends. The remaining three years will be used for testing, this approach will ensure that we are forecasting and calculating our evaluation metrics on unseen, out-of-sample data [11].

C. MLP

MLP means Multilayer Perceptron's. It is the classical type of neural network. It is the set of interconnected neurons. It comprised of one or more layers of neurons. In input layer Data is fed, more than one hidden layers providing levels of abstraction, and predictions are the output layer, also called the visible layer. The activation functions used in such networks are mainly threshold or sigmoid functions.

$$\sigma(y) = \frac{e^y}{e^y+1} = \frac{1}{1+e^{-y}}$$

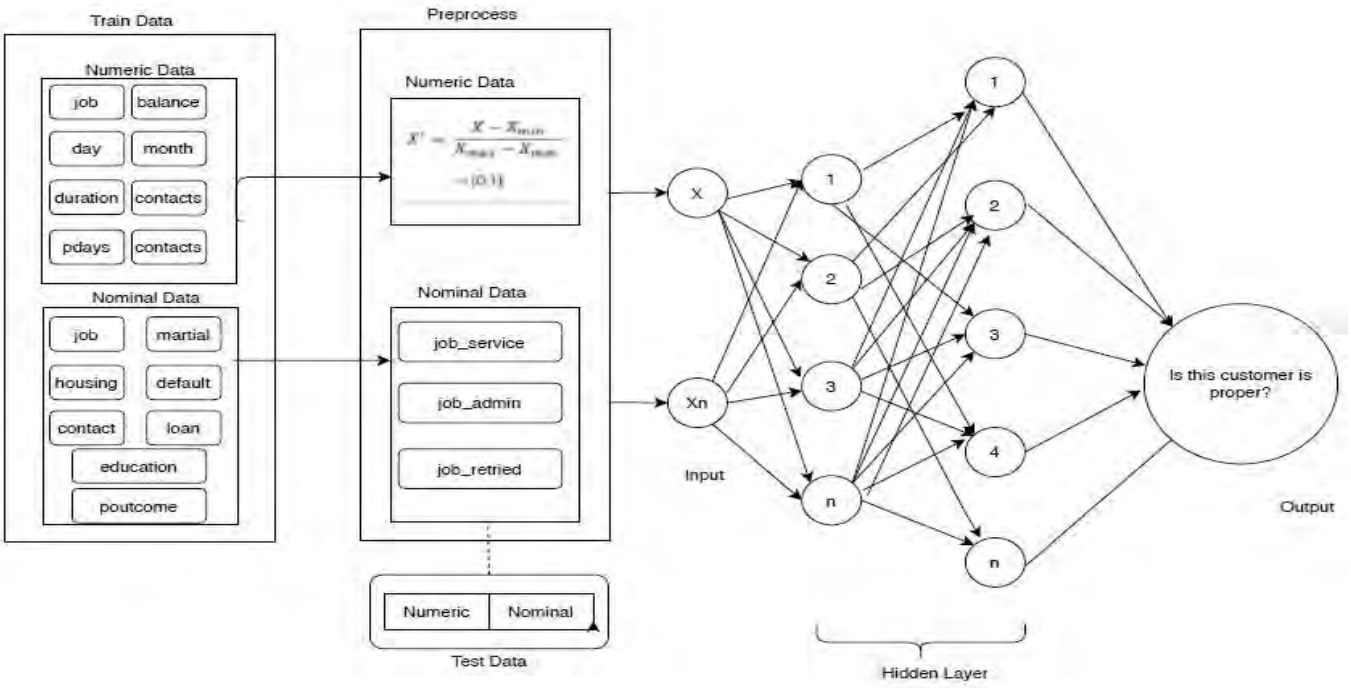


Figure 1: System analysis of bank telemarking using MLP

There are several uses of MLP are:

- Tabular datasets
- Classification prediction problems
- Regression prediction problems[12]

- The input layer: First layer of the network where no calculation is made. The input cell only receives in input the values of attributes: X,… , Xn

- The output layer: Last layer of the network that contains one or more decision cells.[13]

For numerical data: (Ref. Table 1)

- For each numerical feature we calculate directly statistic parameters (mean, max, min, variance, standard deviation).

For categorical data:

- It is distinguish into 3 subtypes features are scaled features, binomial features and nominal features.[14]

For scaled values (Month, days of week):

- For calculating a statistics parameters for numerical values we substitute items by their ordinal number.

For Boolean features:

- For the Boolean features we have two possibilities either yes or no (1/0); else success or failure (1/0).

For Nominal features:

- For each value of the feature Vj for the
- i is replaced by his occurrence frequency which is calculated by:

$$Vij \leftarrow fij = mN \quad (1)$$

Where m is the frequency of the class, and N is the total number of instances.

Having a class which highest in similarity measure (Average of all features) is allocated. [15]

IV. CONCLUSION

In this paper we have proposed a type of ANN which is MLP technology used to predict the direct marketing campaign which is in the main strategy of many banks and insurance companies for interacting with their customers and making new business. We have a technique computes a specific similarity for each type of features. This technique fostered most significant features and predicts the client class more accurately. This paper evaluating the DM technique on the bank direct marketing dataset to classify for bank deposit subscription and increasing the campaign effectiveness by identifying the main characteristics that affect the success

V. ACKNOWLEDGEMENT

We would like to thank “Portuguese bank dataset” for their helpful suggestions.

References

[1]P. Kotler, and K. L. Keller. “Marketing Management”.14th Edition, New Jersey: Prentice Hall, 2012.

[2]https://www.bms.co.in/telemarketing-in-banking- sector/.

[3]https://www.researchgate.net/publication/260805594_ A_DataDriven_Approach_to_Predict_the_Success_of_B ank_Telemarketing.

[4] Y. LeCun *et al.*, “Gradient-based learning applied to document recognition,” IEEE, 1998.

[5] Y. LeCun, “Convolutional neural networks (LeNet), deep learning 0.1 documentation,” University of Montreal, 2013.

[6]Y. Kawasaki and M. Ueki, “Sparse Predictive Modeling for Bank Telemarketing Success Using Smooth- Threshold Estimating Equations,” *J. Jpn. Soc. Comp. Stat.*, vol. 28, pp. 53–66, 2015.

[7]H. A.Elsalamony, “Bank Direct Marketing Analysis of Data Mining Techniques,” *Int. J. Comput. Appl.*, vol. 85, no. 7, pp. 12–22, 2014.

[8]https://www.researchgate.net/publication/330787414_ Optimizing_the_prediction_of_telemarketing_target_call s_by_a_classification_technique.

[9] Y. Tang, *Deep Learning Using Linear Support Vector Machines*, Department of Computer Science, University of Toronto, Toronto, Ontario, Canada, 2013

[10]S. Moro, R. M. S. Laureano and P. Cortez. "Using Data Mining for Bank Direct Marketing: an Application of The CRISP-DM Methodology," in Proceedings of the European Simulation and Modelling Conference -ESM' 2011, Guimarães, 2011.

[11]S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

[12]2017, “Financial Time Series Forecasting -A Deep Learning Approach”.

[13]https://machinelearningmastery.com/when-to-use- mlp-cnn-and-rnn-neural-networks/.

[14] https://publications.waset.org/10010974/predicting- the-success-of-bank-telemarketing-using-artificial- neural-network.

[15] https://archive.ics.uci.edu/ml/datasets/bank+marketin

Comparative Analysis of Different NLP Techniques for Detecting Sarcasm on Various Applications

Priyanka Bolinjkar
M.E. Second Year in
Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
Priyankabolinjkar33@gmail.com

Abstract— **sarcasm is a sophisticated form of irony widely used in social networks and microblogging websites. It is usually used to convey implicit information within the message a person transmits. Sarcasm might be used for different purposes, such as criticism or mockery. However, it is hard even for humans to recognize. Therefore, recognizing sarcastic statements can be very useful to improve automatic sentiment analysis of data collected from microblogging websites or social networks. Sentiment analysis refers to the identification and aggregation of attitudes and opinions expressed by internet users toward a specific topic. We also study the importance of each of the proposed sets of features and evaluate its added value to the classification. In particular, we emphasize the importance of pattern-based features for the detection of sarcastic statements.**

Keywords: pos (part of speech), ner (named entity relation), twitter, sentiment analysis, polarity

I. INTRODUCTION

Sentiment Analysis is a Natural Language Processing (NLP) and data extraction process that helps to acquire user‘s sentiments written in positive, negative, neutral remarks, questions and demands by doing research on large quantities of documents. As a rule, sentiment analysis expects to decide the perception of a speaker or an author as for some point or the general tonality of a report. Sentiments of people can be analyzed as Positive, Negative and Neutral. The sentiment analysis is vital to the business fields. By analyzing, the negative sentiment from the twitter can be used to improve the business and can benefit to the customers. Sarcasm is a kind of sentiment derived from the French word ―Sarkasmos‖ which means ―tear flesh‖ or ―grind the teeth. The meaning is different than what the
speaker intends to say through sarcasm. Sarcasm can also be defined as a ―contrast between a positive sentiment and negative situation‖ and vice versa. Twitter became one of the biggest web destinations for people to express their opinions, share their thoughts and report real-time events, etc. Throughout the previous years, Twitter content continued to increase, thus constituting a typical example of the so-called big data. Today, according to its official website, Twitter has more than 288 million active users, and more than 500 million tweets are sent every day. Many companies and organizations have been interested in these data for the purpose of studying the opinion of people towards political events, popular products or movies . However, due to the informal language used in Twitter and the limitation in terms of characters (i.e., 140 characters per tweet), understanding the opinions of users and performing such analysis is quite difficult. Furthermore, presence of sarcasm makes the task even more challenging: sarcasm is when a person says something different from what he means. Some people are more sarcastic t h a n others, however, in general, sarcasm is very common, though, difficult to recognize. In general, people employ sarcasm in their daily life not only to make jokes and be humorous but also to criticize or make remarks about ideas, persons or events. Therefore, it tends to be widely used in social networks, in particular microblogging websites such as Twitter. The area of Sentiment Analysis intends to comprehend these opinions and distribute them into the categories like positive, negative, neutral. Most sentiment analysis work has been done on review sites . Review sites provide with the sentiments of products or movies, thus, restricting the domain of application to solely business. Sentiment analysis on Twitter posts is the next step in the field of sentiment analysis, as tweets give us a richer and more varied resource of opinions and sentiments that can be about anything from the latest phone they bought, movie they watched, political issues, religious views or the individuals state of mind.

II. II. LITERATURE REVIEW

In the last few years, more attention has been given to Twitter sentiment analysis by researchers, and a number of recent papers have been addressed to the classification of tweets. However, the nature of the classification and the features used vary depending on the aim. Sriram et al. used non-context-related features such as the presence of slangs, time-event phrases, opinioned words, and the Twitter user information to classify tweets into a predefined set of generic classes including events, opinions, deals, and private messages. Akcora et al. proposed a method to identify the emotional pattern and the word pattern in Twitter data to determine the changes in public opinion over the time. They implemented a dynamic scoring function based on Jaccard‘s similarity of two successive intervals of words and

used it to identify the news that led to breakpoints in public opinion.

However, most of the works focused on the content of tweets and were conducted to classify tweets based on the sentiment polarity of the users towards specific topics. A variety of features was proposed. Not only they include the frequency and presence of unigrams, bigrams, adjectives, etc, but they also include non-textual features such as emoticons (i.e., facial expressions such as smile or frown that are formed by typing a sequence of keyboard symbols, and that are usually used to convey the writer‘s sentiment, emotion or intended tone) and slangs . Dong et al. proposed a target-dependent classification framework which learns to propagate the sentiments of words towards the target depending on context and syntactic structure.

Sarcasm, on the other hand, and irony in general have been used by people in their daily conversations for a long time. Therefore, sarcasm has been subject to deep studies form psychological and even neurobiological perspectives. Nevertheless, it has been studied as a linguistic behavior characterizing the human being . In this context, 5478
VOLUME 4, 2016 researchers have recently been interested in sarcasm, trying to find ways to automatically detect it when it is present in a statement. Although some studies such as highlighted that, unlike irony, sarcasm _is not a discrete logical or linguistic phenomenon‘‘, many works
have been proposed and present high accuracy and precision.

Burfoot and Baldwin introduced the task of filtering satirical news articles from true newswire documents. They introduced a set of features including the use of profanity and slangs and what they qualified of _semantic validity‘‘; and used Support Vector Machine (SVM) classifier to recognize satire
articles. Campbell and Katz studied the contextual components utilized to convey sarcastic verbal irony and proposed that sarcasm requires the presence of four entities: allusion to failed expectation, pragmatic insincerity, negative tension and presence of a victim, as well as stylistic components.

III. SCOPE OF NLP

The tweets are differentiated among positive, negative and neutral based on the detection engine. The tweets that are classified as positive, negative and neutral are further classified as Actual positive, Actual negative and Sarcastic sentiment. Actual positive tweets contain the tweets that are actually positive, Actual negative tweets are the tweets that are actually negative and sarcastic tweets are those in which it is difficult to determine whether tweets are actually positive or negative.

Removal of Stop Words:-

The stop words are removed using python programming. The stop words corpus distributed with NLTK have been used. The stop words corpus consist of 2400 stop words for
11 different languages. Example: –This is a long bar sentence‖. After stop words removal we get `long‘ `bar‘sentence‘.
`

Tokenization:

Tokenization is process of breaking texts into words, phrases and symbols called tokens. The tokens generated helps in parsing and performing other mining tasks. Tokenization is done using TextBlob package installed in NLTK. Example: –Beautiful is better than ugly‖. After tokenization using TextBlob we get every word separated
such as `Beautiful‘, `is‘, `better‘, `than‘, `ugly‘.

Part of Speech Tagging :

Part of Speech Tagging divides sentences into words and assign part of speech for each word. In this paper TextBlob is used for Part of Speech Tagging. Example: –Sarcasm detection using twitter data‖. Part of Speech tagging for the example is Sarcasm-NNP, detection-NN, using-VBG, twitter-NN, data- NNS. Here NNP stands for proper noun singular, NN stands for noun singular, VBG stands for verb present participle and NNS stands for noun plural.

III. APPROACHES AND COMPONENTS OF SARCASM DETECTION

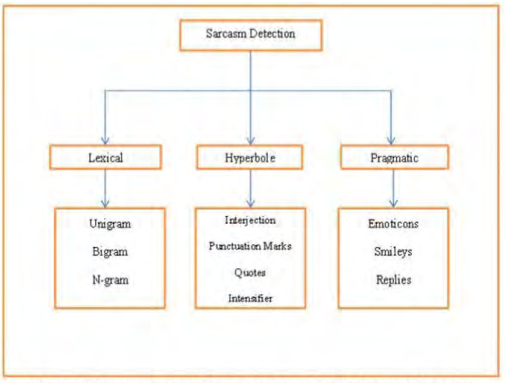


Figure [1] Approaches and Components used in the Sarcasm Detection

Lexical analysis: It the process of analyzing a stream of individual characters, into a sequence of lexical tokens to feed into the parser. Performs reading of input characters, removal of white spaces and comments. Roughly the equivalent of splitting ordinary text written in a natural language (e.g. English) into a sequence of words and punctuation symbols. Strictly speaking, tokenization may be handled by the parser. The reason why we tend to bother with tokenising in practice is that it makes the parser simpler, and decouples it from the character encoding used for the source code.

Hyperbole: It is a device that we employ in our day-to- day speech. For instance, when you meet a friend after a long time, you say, –It‘s been ages

since I last saw you.‖ You may not have met him for three or four hours, or a day, but the use of the word –ages‖ exaggerates this statement to add emphasis to your wait.

Examples of hyperbole:

1. The blacksmith‘s hand was harder than the rock.
2. The old man was older than the Himalayas.

Pragmatic: Pragmatics is a subfield of linguistics and semiotics that studies the ways in which context contributes to meaning. Pragmatics encompasses speech act theory, conversational implicative, talk in interaction and other approaches to language behavior in philosophy, sociology, linguistics and anthropology.
Various Approaches used in Sarcasm detection

1. Rule Based Approach: Rule-based system is used to store and manipulate knowledge to interpret information in a useful way. It is often used in artificial intelligence applications and research. A classic example of a rule-based system is the domain-specific expert system that uses rules to make deductions or choices. For example, an expert system might help a doctor choose the correct diagnosis based on a cluster of symptoms, or select tactical moves to play a game. Rule- based systems can be used to perform lexical analysis in natural language programming.
2. Supervised Approach : Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from *labeled training data* consisting of a set of *training examples*. In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value (also called the *supervisory signal*). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples

.

Following are some key characteristics of various applications:

1. Message Length: The maximum length of a Twitter message is 140 characters. This is different from previous sentiment classification research that focused on classifying longer texts, such as product and movie reviews.

2. Writing technique: The occurrence of incorrect spellings and cyber slang in tweets is more often in comparison with other domains. As the messages are quick and short, people use acronyms, misspell, and use emoticons and other characters that convey special meanings.

3. Availability: The amount of data available is immense. More people tweet in the public domain as compared to Facebook (as Facebook has many privacy settings) thus making data more readily available. The Twitter API facilitates collection of tweets for training.

V. DESIGN AND ANALYSIS

Tools used

1. Python and visual studio code
2. Libraries

- Textblob :To perform Sentiment Analysis
- Tweepy: To fetch twitter data
- Tweeter Developer: It is an API used to fetch keys

• Def clean_tweet: Used for cleaning of tweets
• Ptweets: positive tweets

• Ntweets: negative tweets

V. METHODOLOGY

The tweets are differentiated among positive, negative and neutral based on the detection engine. The tweets that are classified as positive, negative and neutral are further classified as Actual positive, Actual negative and Sarcastic sentiment. Actual positive tweets contain the tweets that are actually positive, Actual negative tweets are the tweets that are actually negative and Sarcastic tweets are those in which it is difficult to determine whether tweets are actually positive or negative.

Removal of Stop Words:-
The stop words are removed using python programming. The stop words corpus distributed with NLTK have been used. The stop words corpus consist of 2400 stop words for
11 different languages. Example: –This is a long bar sentence‖. After stop words removal we get `long‘ `bar‘ `sentence‘.

Tokenization
Tokenization is process of breaking texts into words, phrases and symbols called tokens. The tokens generated helps in parsing and performing other mining tasks. Tokenization is done using TextBlob package installed in NLTK. Example: –Beautiful is better than ugly‖. After tokenization using TextBlob we get every word separated such as `Beautiful‘, `is‘, `better‘, `than‘, `ugly‘.

Part of Speech Tagging
Part of Speech Tagging divides sentences into words and assign part of speech for each word. In this paper TextBlob is used for Part of Speech Tagging. Example: –Sarcasm detection using twitter data‖. Part of Speech tagging for the example is Sarcasm-NNP, detection-NN, using-VBG, twitter-NN, data-NNS. Here NNP stands for proper noun singular, NN stands for noun singular, VBG stands for verb present participle and NNS stands for noun plural.

Parsing
Parsing is process of analyzing the symbols based on natural

language. When a sentence is passed through parser, the parser divides the whole sentence into words and finds Indian Journal of Science and Technology Vol 10 (25) | July 2017 | www.indjst.org 5 Shubhodip Saha, Jainath Yadav and Prabhat Ranjan Part of Speech (POS) tags.POS and syntactic relations are used to form units such as subject, verb and object. In this paper python based package named TextBlob is used for parsing. Example: ‒Sarcasm detection using twitter data‖ is parsed giving result as Sarcasm/NN/B- NP/O, detection/ NN/I-NP/O, using/VBG/BVP/O,twitter/NNP/B-NP/O, data/NNS/I-NP/O. Here O represents out

Sentiment Related Features.

A very popular type of sarcasm that is widely used in both regular conversations as well as short messages such as tweets, is when an emotionally positive expression is used in a negative context. A similar way to express sarcasm is to use expressions having contradictory sentiments. This type of sarcasm we qualified as ‗whimper‘‘ is very common in social networks and microblogging websites. Riloff et al. show that this type of sarcasm can be identified and detected when a positive statement, usually a verb or a phrasal verb, is collocated with a negative situation (e.g., ‗I love being ignored all the time‘‘). They built a lexicon-based approach that learns the possible positive expressions and negative situations and used it to detect such contrast in unknown tweets. Therefore, to identify and quantify such inconsistency, we extract sentimental components of the tweet and count them. For this purpose, we maintain two lists of words qualified as ‗positive words‘‘ and ‗negative words‘‘. The two lists contain respectively words that have positive emotional content (e.g., ‗love‘‘, ‗happy‘‘, etc.) and negative emotional content (e.g., ‗hate‘‘, ‗sad‘‘, etc.). The two lists of words are created using SentiStrength6 database. This database contains a list of emotional words, where negative words have scores varying from −1 (almost negative) to −5 (extremely negative) and positive words have score varying from 1 (almost positive) to 5 (extremely positive). Using these two lists, we extract two features we

denote respectively pw and nw by counting the number of positive and negative words in the tweet.

IV. COMPARISON OF DIFFERENT NLP TECHNIQUES

SVM	Hybrid	NER	Text Summarization
Support machine is a kind of large machine classifier: here the goal I is to find a decision boundary between 2 classes	It is used for text clustering which is the combination of active feature selection genetic algorithm and bisecting k- Gram	It is subtask of information extraction that seeks to locate and classify name entity mentioned in unstructured text into predefined categories like name,person,v	It is a process of shortening a set of data computantion ally,to create a subset of that represents the must important or relevant information within the original
Many NLP problems can be formulated as classification Eg.Word sense disabmiguation,spam Filtering,Etc	Improving NLP Application. If We are developing grammar correction system, Rule based module that identifies wrong pattern and generate tight pattern.	Most research on NER system has been as taking an unnoted block text Eg. Jim Bought 300 shares of Acme corp. organization in 2006 [Jim] –Person [Acme Corp]- [2006]-time	In addition to text, images and videos can also be summarized. Eg. key pharses that can be used to ‒tagllor index a text document
More effective in wgh dimension spaces	Much better suited to classification with complex non-distributed classes.	It provides different models for pulling out name entities.	Works in any language, improves productivitty

V. CONCLUSION

Using the senti-strength tool we get the polarity of each word as well as the overall sentence. For the sentence, a polarity score of -1 to -5 is considered to be negative polarity sentence, 0 is taken to be neutral while a score of +1 to +5 is taken to be of positive polarity.

VI. ACKNOWLEDGEMENT

We hereby take the privileged to present our paper on NLP Techniques for detecting Scram on Twitter . We are very grateful to our paper guide Dean Dr.R.R.Sedamkar for contributing and valuable time in the paper from their busy and hectic schedule. Thank you for being after us like a true mentor and great academic parents.

We are very thankful to Dean Dr.R.R.Sedamkar whose guidance and support was an immense motivation for us to carry on with our paper. Also suggestions have greatly contributed for the betterment of our paper.

VII. References

[1] Jehoshua Eliashberg, Quintus Hegie , Jason Ho, Dennis Huisman , ―Scheduling of movie in a multiplex,‖,2009.

[2] Antara Upadhyay, Nivedita Kamath, Shalin Shanghavi, Tanisha Mandvikar, Pranali Wagh ―Movie Success Prediction using Data Mining,‖ in IJEDR, vol. VI,Issue IV,Shah & Anchor Kutchhi Engineering College, Mumbai:Academic, pp. 198–203

[3] Jim Hee Jinny Lim, ― A Thesis on Forecasting Movie Attendance of Individual Movie Showings: A Hierarchical Bayes Approach‖,Simon Fraser University,2012.

[4] Dursun Delen and Ramesh Sharda ―Artificial Neural Networks in Decision Support Systems‖,Spears School of Business Department of Management Science andInformation Systems ,Oklahoma State University, January 2008.

[5] N. Quader, Md. O. Gani, D. Chaki, Md. H Ali, ―A Machine Learning Approach to Predict Movie Box-Office Success‖, in 20th International Conference of Computer and Information Technology -(ICCIT), 22-24 December, 2017.

[6] Javaria Ahmad,Prakash Duraisamy,Amr Yousef,BillBuckles, ―Movie Success Prediction Using Data Mining‖, in 8th International Conference On Computing,Communication And Networking Technologies (ICCCNT),3 -5 July, 2017.

[7]Edward Benson, Aria Haghighi, and Regina Barzilay. 2011. Event discovery in social media feeds. In The 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon, USA. To appear.

[8] David M. Blei, Andrew Y. Ng, and Michael I. Jordan.2003. Latent dirichlet allocation. J. Mach. Learn. Res.

[9] Avrim Blum and Tom M. Mitchell. 1998. Combining labeled and unlabeled sata with co-training. In COLT, pages 92–100.

[10]Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai. 1992. Classbased n-gram models of natural language. Comput. Linguist.

The Study of Ensemble's for Recommending Better Future Finance

Kajal Bholashankar Jaiswal, Dr. Harshali Patil
Dept. Of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
kajalj2222@gmail.com, harshali.patil@thakureducation.org

Abstract— **As a rule, House value record speaks to the condensed value changes of private lodging. While at a solitary family house cost expectation, it needs increasingly exact technique dependent on area, house type, size, form year, nearby civilities, and some different components which could influence house request and supply. With restricted dataset and information includes, a pragmatic and composite information pre-handling, innovative element building strategy is inspected right now.**

Keywords: Regression, Lasso, Ridge, Ada-boost, city, comfort, investments, house price, estate agent, property

I. INTRODUCTION

Land Property isn't just the essential need of a man yet today it additionally speaks to the wealth and notoriety of an individual. Interest in land by and large is by all accounts productive in light of the fact that their property estimations don't decrease quickly. Changes in the land cost can influence different family unit financial specialists, brokers, approach creators and many. Interest in land segment is by all accounts an alluring decision for the ventures. Accordingly, foreseeing the land esteem is a significant monetary record. India positions second on the planet in number of family units as indicated by 2011 evaluation with various 24.67 crore. India is additionally the quickest developing significant economy in front of China with previous' development rate as 7% this year and anticipated to be 7.2% in the following year.

According to the 2017 version of Emerging Trends in Real Estate Asia Pacific, Mumbai and Bangalore are the topranked cities for investment and development. These cities have supplanted Tokyo and Sydney. The house prices of 22 cities out of 26 dropped in the quarter from April to June when compared to the quarter January to March according to National Housing Bank's Residex(residential index). With the introduction of Real Estate Regulation Development Act (RERA) and Benami property Act throughout the country India, more number of investors are attracted to invest into real estate in India. The strengthening and modernizing of the Indian economy has made India as attractive Investment destination. However, past recessions show that real estate prices cannot necessarily grow. Prices of the real estate property are related to the economic conditions of the state [1]. Despite this, we are not having proper standardized ways to measure the real estate property values.

Generally the property values rise with respect to time and its appraised value need to be calculated. This appraised value is required during the sale of property or while applying for the loan and for the marketability of the property. These appraised values are determined by the professional appraisers. However, drawback of this practice is that these appraisers could be biased due to bestowed interests from buyers, sellers or mortgages. Thus, we require an automated prediction model that can help to predict the property values without any bias. This automated model can help the first time buyers and less experienced customers to understand whether the property rates are overrated or underrated.

Now, Property prices depend on various parameters in the economy and society. However, previous analyses show that house prices are strongly dependent on the size of the house and its geographical location [2], [3]. We have also considered various intrinsic parameters (such as number of bedrooms, living area and construction material) and also external parameters (such as location, proximity, upcoming projects, etc.) [4], [5]. Then we have applied these parameter values to two different machine learning algorithms.

II. Literature Review

In most recent two decades anticipating the property estimation has become a significant field. Ascend in the interest for property and unusual conduct of economy urge analysts to discover a way that anticipate the land costs with no inclinations. Therefore, it is a test for scientists to discover all the moment factors that can influence the expense of property and make a prescient model by thinking about all the variables. Building a prescient model for land value valuation requires a careful information regarding the matter. Numerous

specialists have dealt with this issue and imparted their examination work.

The majority of this exploration work is enlivened from [6]. The creator has scratched the lodging informational index from Centris.ca and duProprio.com. Their dataset comprises of around 25,000 models and 130 components. Around 70 highlights were scratched from the above sites and land offices, for example, RE/MAX, Century 21, and Sutton, and so forth. Other 60 highlights were sociodemographic dependent on where the property is found. Afterward, creator actualized Principal Component Analysis to diminish the dimensionality. The creator utilized four relapse strategies to anticipate the value estimation of the property. The four

procedures are Linear Regression, Support Vector Machine, KNearest Neighbors (KNN) and Random Forest Regression and an outfit approach by consolidating KNN and Random Forest Technique. The gathering approach anticipated the costs with least blunder of 0.0985. Notwithstanding, applying PCA didn't improve the forecast mistake.

A lot of researches have been done on Artificial Neural Networks. This has helped many researchers focusing on real estate problem to solve using neural networks. In [7], the author has compared hedonic price model and ANN model that predict the house prices. Hedonic price models are basically used to calculate the price of any commodity that are dependent on internal characteristics as well as external characteristics. The hedonic model basically involves regression technique that considers various parameters such as area of the property, age, number of bedrooms and so on. The Neural Network is trained initially and the weights and biases of the edges and nodes respectively are considered using trial and error method. Training the Neural Network model is a black box method. However, the RSquared value for Neural Network model was greater compared to hedonic model and the RMSE value of Neural Network model was relatively lower. Hence it is concluded that Artificial Neural Network performs superior than Hedonic model.

Some researchers like that in [8] have used classifiers to predict the property values. The author in research article [8] has collected the data from Multiple Listing Service (MLS), historical mortgages rates and public school ratings. Real Estate Data was obtained from Metropolitan Regional Information Systems (MRIS) database. The author extracted approximately 15,000 records from these three sources which included 76 variables. Subsequently, t-test was used to select 49 variables as a preliminary screening.

Their research question was to determine whether the closing price was higher or lower than the listing price [8]. Thus to address this classification problem, the author used four machine learning models. C4.5, RIPPER, Naive Bayesian, and AdaBoost are the four algorithms used by author. However, they found that RIPPER outperforms other house prediction models. However the drawback is that performance evaluation is based only on classifiers. Performance comparison of other machine learning algorithms should also be considered.

In article [9], the authors have predicted the stock market prices using linear regression technique. They have collected stock market data from TCS stock Database. The author have also used RBF and polynomial regression technique along with linear regression and found that latter is superior to the remaining techniques.

In [11], the author has considered the most macroeconomic parameters that affect the house prices variation. In this, the author has used back propagation neural network (BPN) and radial basis function neural network (RBF) to establish the nonlinear model for real estates price variation prediction. The dataset have been taken from Taipei, Taiwan based on leading and simultaneous economic indices. The author has considered 11 parameters. The prediction results obtained from them are compared to public Cathay House Price Index or the Sinyi Home Price Index. The two error metrics used were Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). When the prediction results were compared to Cathay House Price Index, RBF Neural Network showed better prediction results than BPN Neural Network. Similarly, for Sinyi Home Price Index BPN Neural Network showed better prediction results than RBF Neural Network.

Some examination articles portray the inside and out strategies and methods to gather the land information and their preprocessing procedures. The writer in article [12] portrays programming that is utilized in land value assessment. The product examines different land servers and site pages of land organizations, and records their present connects to land buy or rental into their product database. He has accumulated information from Czech Republic. The information is accumulated each month to record the progressions occurring in land. The product assembles 110,000 sections each month. These passages incorporate different writings, notices and pictures of the property. The creator has gathered information from the year 2007 to 2015.This unstructured information that is gathered is traded into an organized structure. Diverse property types have various parameters. In this manner it makes the informational collection increasingly clear. This informational index is then

assessed. New passages made every month are contrasted with the more seasoned sections and checked for their culmination. In the last period of the product, this accessible clean informational collection is then assessed and delivers different perceptions as per the necessity of the client. In this manner the yield got might be utilized as justification for fitting speculations as well as lodging choices for both normal people and organizations.

Some researchers have focused on feature selection and feature extraction procedure. The author in article [14] uses a open source data set of the housing sales in King County, USA. There are about 20 explanatory variables. The author has compared various feature selection and feature extraction algorithms combined with Support Vector Regression. The author has collected approximately 21,000 observations in a time period of one year. The paper shows various data analysis performed on the data set. Feature Selection is the process of selecting a subset of variables from a given set of parameters either based on their importance or their frequency. However, feature extraction is the process of reducing the dimensionality of the data. Initial set of data is transformed into derived values which are equally informative and non-redundant. The three feature selection algorithms used are Recursive Feature Elimination (RFE), Lasso and Ridge and Random Forest Selector and the mean from each algorithm is calculated. Using feature selection, the author selects fifteen features out of twenty. The feature extraction algorithm used is Principal Component Analysis (PCA) and reduces the parameters from twenty to sixteen. However, the author found that both the techniques work equally well with the R squared value of 0.86.

III. COMPARATIVE ANALYSIS

Algorithm	Hyperparameters
AdaBoost (AB)	**n-estimators:** Number of decision trees in the ensemble. **Learning rate :** Shrinks the contribution of each successive decision tree in the ensemble.
Gradient Tree Boosting(GTB)	**n-estimators** - Number of decision trees in the ensemble. **Learning rate :** Shrinks the contribution of each successive decision tree in the ensemble. **Loss:** Loss function to be optimized via gradient boosting. **Max_dept**: Maximum dept of the dept of the decision trees. Controls the complexity of decision tree **Max_features:** Number of features to consider when computing the best node split.
Random Forest (RF)	**n-estimators** - Number of decision trees in the ensemble. **Max_features:** Number of features to consider when computing the best node split. **Criterion:** Function used to measure the quality of a split. **Min_weight-fraction_leaf:** Number of features to consider when computing the best node split.
Decision Tree (DT)	**Max_features:** Number of features to consider when computing the best node split. **Criterion:** Function used to measure the quality of a split. **Min_weight-fraction_leaf:** Number of features to consider when computing the best node split.
K-Nearest neighbour (KNN)	**n_neighbours** – Number of neighbours to use. **weights**: Function to weight the neighbour votes.
Support Vector Classifier (SVC)	**kernel**: 'linear', 'poly', 'sigmoid', or 'rbf' **C :** Penalty parameter for regularization. **Gamma:** Kernel coefficient for 'rbf', 'poly' and 'sigmoid' kernels **Degree:** degree for the 'poly' kernel.
Logistic Regression (LR)	**C:** Regularization of strength. **Penalty**: whether to use lasso or ridge regularization. **Fit_intercept** : whether or not the intercept of the linear classifier should be computed.
Linear Regression	The basic form where the model is penalized for its choices of weights. That means, during the training stage, if the model feels like one particular feature is particularly important, the model may place a large weight to the feature. This sometimes leads to overfitting in small datasets. Hence, following methods are invented.
Lasso	Lasso is a modification of linear regression, where the model is penalized for the sum of absolute values of the weights. During training, the objective function become: $\frac{1}{2m}\sum_{i=1}^{n}(y_i - Xw_i)^2 + alpha\sum_{i=1}^{n}\lvert w_i\rvert$ Lasso introduced a new hyperparameter, alpha, the coefficient to penalize weights.
Ridge	Ridge takes a step further and penalize the model for the sum of squared value of the weights. Thus, the weights not only tend to have smaller absolute values and more evenly distributed, but also tend to be close to zeros. The objective function becomes: $\sum_{i=1}^{n}(y_i - Xw_i)^2 + alpha\sum_{i=1}^{n}w_i^2$
ElasticNet	ElasticNet is a hybrid of Lasso and Ridge, where both the absolute value penalization and squared penalization are included, being regulated with another coefficient l1_ratio:

Fig 2 : Comparative Analysis of Regression Algorithms

IV. CONCLUSION

In this research paper, we have used machine learning algorithms to predict the house prices. We have mentioned the step by step procedure to analyze the dataset and finding the correlation between the parameters. Thus we can select the parameters which are not correlated to each other and are independent in nature. These feature set were then given as an input to four algorithms and a csv file was generated consisting of predicted house prices. Hence we calculated

the performance of each model using different performance metrics and compared them based on these metrics.

For future work, we recommend that working on large dataset would yield a better and real picture about the model. We have undertaken only few Machine Learning algorithms that are actually classifiers but we need to train many other classifiers and understand their predicting behavior for continuous values too. By improving the error values this research work can be useful for development of applications for various respective cities.

V. References

[1] R. J. Shiller, "Understanding recent trends in house prices and home ownership," National Bureau of Economic Research, Working Paper 13553, Oct. 2007.

[2] D. Belsley, E. Kuh, and R. Welsch, Regression Diagnostics: Identifying Influential Data and Source of Collinearity. New York: John Wiley, 1980.

[3] J. R. Quinlan, "Combining instance-based and modelbased learning," Morgan Kaufmann, 1993, pp. 236–243.

[4] S. C. Bourassa, E. Cantoni, and M. Hoesli, "Predicting house prices with spatial dependence: a comparison of alternative methods," Journal of Real Estate Research, vol. 32, no. 2, pp. 139–160, 2010.[Online]Available:http://EconPapers.repec.or g/RePEc:jre:issued:v:3 2: n:2:2010:p:139-160.

[5] S. C. Bourassa, E. Cantoni, and M. E. Hoesli, "Spatial dependence, housing submarkets and house price prediction," eng, 330; 332/658, 2007, ID: unige:5737.[Online]. Available: http:// archive - ouverte. unige. ch/unige:5737.

[6] Pow, Nissan, Emil Janulewicz, and L. Liu. "Applied Machine Learning Project 4 Prediction of real estate property prices in Montréal." (2014).

[7] Limsombunchai, Visit. "House price prediction: hedonic price model vs. artificial neural network."New Zealand Agricultural and Resource Economics Society Conference. 2004.

[8] Park, Byeonghwa, and Jae Kwon Bae. "Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data."Expert Systems with Applications 42.6 (2015): 2928-2934.

[9] Bhuriya, Dinesh, et al. "Stock market predication using a linear regression." Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of.Vol. 2.IEEE, 2017.

[10] Majumder, Manna, and MD Anwar Hussian. "Forecasting of Indian stock market index using artificial neural network."Information Science (2007): 98-105.

[11] Li, Li, and Kai-Hsuan Chu. "Prediction of real estate price variation based on economic parameters." Applied System Innovation (ICASI), 2017 International Conference on.IEEE, 2017.

[12] Hromada, Eduard. "Mapping of real estate prices using data mining techniques." Procedia Engineering 123 (2015): 233- 240.

[13] Y. P. Anggodo, A. K. Ariyani, M. K. Ardi, and W. F. Mahmudy,―Optimation of Multi-Trip Vehicle Routing Problem with Time Windows using Genetic Algorithm,‖ J. Environ. Eng. Sustain. Technol., vol. 3, no. 2, pp. 92–97, 2017.

[14] R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartotoor, and I. B. Syamwil,Factors influencing the price of housing in Indonesia,‖ Int. J. Hous.Mark. Anal., vol. 8, no. 2, pp. 169– 188, 2015.

[15] V. Limsombunchai, ―House price prediction: Hedonic price model vs.artificial neural network,‖ Am. J. …, 2004.

[16] D. X. Zhu and K. L. Wei, ―The Land Prices and Housing Prices —— Empirical Research Based on Panel Data of 11 Provinces and Municipalities in Eastern China,‖ Int. Conf. Manag. Sci. Eng., no. 2009,pp. 2118–2123, 2013.

[17] S. Kisilevich, D. Keim, and L. Rokach, ―A GIS-based decision support system for hotel room rate estimation and temporal price prediction: The hotel brokers' context,‖ Decis. Support Syst., vol. 54, no. 2, pp. 1119–1133, 2013. [18] C. Y. Jim and W. Y. Chen, ―Value of scenic views: Hedonic assessment of private housing in Hong Kong,‖ Landsc. Urban Plan., vol. 91, no. 4, pp. 226–234, 2009.

[18] L. Bryant, ―Housing affordability in Australia: an empirical study of the impact of infrastructure charges,‖ J. Hous. Built Environ., 2016.

[19] S. Rosen, ―Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,‖ J. Polit. Econ., vol. 82, no. 1, pp. 34–55, 1974.

Diabetic Retinopathy: A General Survey

Nidhi Kamothi, Rashmi Thakur
Thakur College of Engineering & Technology
University Of Mumbai
Mumbai, India
nidhikamothi95@gmail.com, thakurrashmik@gmail.com

Abstract— **Diabetic retinopathy (DR) is one of the leading causes of vision loss and can be effectively avoided by screening, early diagnosis and treatment. The retina is the only part of the human body where the blood circulation can be observed directly. Diabetic retinopathy is caused by the retinal micro vasculature which may be formed as a result of diabetes mellitus. Blindness may appear as a result of unchecked and severe cases of diabetic retinopathy.**

Manual inspection of fundus images to check morphological changes in micro aneurysms, exudates, blood vessels, hemorrhages, and macula is a very time-consuming and tedious work. It can be made easily with the help of computer-aided system and intervariability for the observer. In this paper, we provide overview of a computer-based approach for the detection of diabetic retinopathy using retinal images. There are many features present in retina but the exudates feature which is one of the primary signs of diabetic retinopathy and which is a main cause of blindness that could be prevented with the help of this automatic detection process technique.

The automatic detection process reduces examination time, and increase accuracy. In this paper provide review on many techniques and algorithms that helps to diagnose Diabetic Retinopathy in retinal fundus images.

Keywords— Diabetic Retinopathy, Fundus images, Exudates, Retinal blood vessels, macula, microaneurysms.

I. INTRODUCTION

Diabetes is a very common disease worldwide. It serves as a most common cause of blindness for people having age less than 50 years. Diabetes is the chronic state caused by an abnormal increase in the glucose level in the blood and which causes the damage to the blood vessels. The tiny blood vessels that nourish the retina are damaged by the increased glucose level. It is a systemic disease which is affecting up to 80 percent of people for more than 10 years. Many researchers acknowledged that 90 percent of diabetic patients could be saved from this disease through an early diagnose. A person having diabetes is more prone to the risk of diabetic retinopathy (DR) [1].

Diabetic retinopathy (DR) occurs when diabetes damages the tiny blood vessels inside the retina, the light- sensitive tissue at the back of the eye [2]. This tiny blood vessel will leak blood and fluid on the retina, forming features such as microaneurysms (MAs), haemorrhages (HMs), hard exudates (EXs). Microaneurysms are an earlier sign of DR. DR can be broadly classified as non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) [3]. Depending on the presence of features on the retina, the stages of DR can be identified [3]. DR has four stages [3, 4].

1. Mild non-proliferative retinopathy. At least one micro aneurysm with or without the presence of retinal haemorrhages, hard exudates, cotton wool spots, or venous loops will be present. Approximately 40 per cent of people with diabetes has at least mild signs of DR [5].
2. Moderate non-proliferative retinopathy. Numerous microaneurysms and retinal haemorrhages will be present. Cotton wool spots and a limited amount of venous beading can also be seen. 16 per cent of the patients with moderate NPDR will develop PDR within 1 year [6].

3. Severe non-proliferative retinopathy. This is characterized by any one of the following characteristics:

(a) numerous haemorrhages and microaneurysms in four quadrants of the retina;
(b) venous beading in two or more quadrants;
(c) intraregional micro vascular abnormalities in at least one quadrant;

4. Proliferative retinopathy. This is the advanced the signals sent by the retina for nourishment trigger the growth of new blood vessels. These new blood vessels are abnormal and fragile.

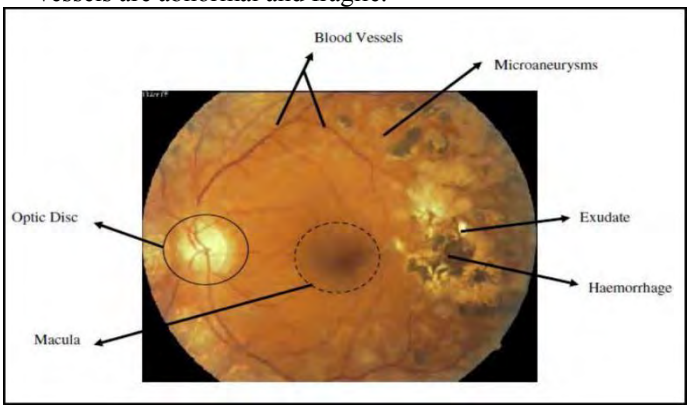


Figure 1: Features in DR image

II. LITERATURE SURVEY

Survey of various research papers are done in this section.

Sr	Title	Author & year of publication	Key findings	Research Gap
01	A brief review of the Detection of Diabetic Retinopathy in Human Eyes Using Pre-Processing & Segmentation Techniques	Yogesh Kumaran, Chandrashekar M. Patil , December 2018	Gives a brief insight into the detection of DR in human eyes It uses different types of preprocessing & segmentation techniques.	processing time is not taken into the account
02	A Survey Paper on Diabetes Retinopathy	Dr. Ruksar Fatima, Laxmi Math, Sept 2017	presents a fast, effective and robust way of detecting diabetic features in the fundus images. Various classification schemes have been applied for retinopathy	the processing time can be decreased in future.
03	Detection of Diabetic Retinopathy: A Survey	Nisha A. Panchal, Dr. Darshak G. Thakore, April 2018	presents a fast, effective and robust way of detecting diabetic features in the fundus images.	processing time is not taken into the account.
04	A Survey on methods of automatic detection of Diabetic Retinopathy	Swati Gupta, A.M. Karandikar, January 2015	Provides overview of a computer-based approach for the detection of diabetic retinopathy using retinal images. The automatic detection process reduces examination time, and increase accuracy.	performance of the proposed model is satisfactory.
05	Automatic detection and classification of diabetic retinopathy stages using CNN	Ratul Ghosh, Kuntal Ghosh, Sanjit Maitra, 2017	A model is presented for classification of DR stages based on the severity using color fundus images.	Proper accuracy is not achieved.

III. PROBLEM DEFINITION

Diabetic Retinopathy is a condition that occurs as a result of damage to the blood vessels of the retina in people who have diabetes. Diabetic retinopathy can develop if you have type 1 or 2 diabetes and a long history of uncontrolled high blood sugar levels. While you may start out with only mild vision problems, you can eventually lose your sight. Untreated diabetic retinopathy is one of the most common causes of blindness in the United States, according to the National Eye Institute. It's also the most common eye disease in people with diabetes. There are various factors affecting the disease like age of diabetes, poor control, pregnancy but Researches shows that progression to vision impairment can be slowed or averted if DR is detected in early stage of the disease. There is a need to detect DR at early levels, so that the damage to the retina can be minimized as it is a common persistent problem faced up to date.

IV. DESCRIPTION OF CNN

Convolutional networks (ConvNets) have recently enjoyed a great success in large-scale image and video recognition which has become possible due to the large public image repositories, such as ImageNet (Deng et al., 2009), and high-performance computing systems, such as GPUs or large-scale distributed clusters

A Convolutional Neural Network (CNN) is comprised of one or more convolutional layers (often with sub-sampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. The architecture of a CNN is designed to take advantage of the 2D hierarchical structure of an input image (or other 2D input such as a speech signal).

Another benefit of CNNs is that they are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units. CNNs also consider the hierarchical representation of images while training by stacking multiple trainable stages on each other.

V. METHODOLOGY AND STRATEGY

There are various methods used in process of detection of DR.
Divide the process into three steps pre-processing, the feature of DR detection and classification.

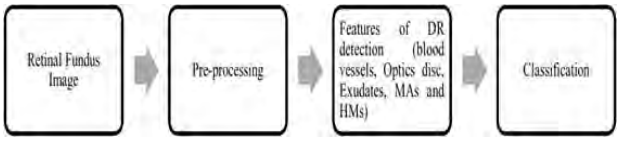


Figure 2: Flowchart of Diabetic Retinopathy

A. Retinal Fundus ImagePre-processing

Pre-processing is an important step in process of detecting DR. Pre-processing is required to ensure that the dataset is consistent and displays only relevant features. Next, the images are segmented to differentiate between the normal and abnormal substances.

B. Feature Extraction

In this step, various features of DR retinopathy are extracted.

Blood Vessels detection: MAs, HMs, and EXs are the main reason for blood vessel damage. So the blood vessels detection is an important factor for detection of DR.

Optic disc detection: The Optic disc is a bright yellow portion and approximately circular in shape. To detect EXs, we need to detect a portion of the optic disc in retinal fundus image because Optic disc and EXs both have same intensity.

EXs Detection: Exudates are bright yellow objects in retinal fundus image which have less than 100 pixels in fundus image. The main objective of exudate detection is the removal of the optic disc before the onset of the process.

MAs and HMs Detection: Microaneurysms and Haemorrhages are first DR symptoms. If we detect the earlier stage then we can reduce the blindness causes from DR. Two methods are used for detecting MAs and HMs. The first method is Semi-automated DR detection using Eigenvalue of Hessian matrix Analysis. The second method is automatic DR detection using Eigenvalue of hessian matrix analysis, image processing techniques, and SVM classifier.

C. Classification

The data is taken from Messidor and then preprocessed using the process of image segmentation. Once that is done, the image is then selected and classified using SVM classifier.

VI. DATASET

We have taken the dataset from the KAGGLE coding website.

VII. EXPECTED OUTCOME

Our proposed model expects the following outcomes:

a. Diabetic retinopathy is best diagnosed with a comprehensive dilated eye exam.
b. Visual loss from diabetic retinopathy is both preventable and treatable.
c. Detection of early, pre-symptomatic disease allows timely treatment with laser photocoagulation to prevent visual loss.
d. The trained CNN makes a quick diagnosis and instant response to a patient possible.

VIII. CONCLUSION

Different types of classifier and its performance are analysed for the automated diagnosis of diabetic retinopathy from the features extracted. There are several detection algorithms that have already been developed and proposed which perform satisfactorily.

Our study has shown that the five-class problem for national screening of DR can be approached using a CNN method. Our network has shown promising signs of being able to learn the features required to classify the fundus images, accurately classifying the majority of proliferative cases and cases with no DR. Our method produces comparable results to these previous methods without any feature-specific detection and using a much more general dataset.

The potential benefit of using our trained CNN is that it can classify thousands of images every minute allowing it to be used in real-time whenever a new image is acquired. The trained CNN makes a quick diagnosis and instant response to a patient possible. This survey paper can act as a resource for the future researchers interested in automated detection of abnormal signs of diabetic retinopathy and help them to get an overview of this field in order to develop more efficient algorithms.

IX. ----

[1] Yogesh Kumaran, Chandrashekar M. Patil, “A Brief Review of the Detection of Diabetic Retinopathy in Human Eyes Using Pre-Processing & Segmentation Techniques”, International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-4S2, December 2018

[2] Dr. Ruksar Fatima, Laxmi Math, “A Survey Paper on Diabetes Retinopathy”, International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), Vol 4, Issue 9, September 2017

[3] Nisha A. Panchal, Dr. Darshak G. Thakore, Dr. Tanmay D. Pawar, “Detection of Diabetic Retinopathy: A Survey”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6, Issue IV, April 2018

[4] Swati Gupta, A.M. Karandikar, “ A Survey on methods of automatic detection of Diabetic Retinopathy,” International Journal of Research in IT, Management and Engineering, Volume 5, Issue 1,

January 2015

[5] Ratul Ghosh, Kuntal Ghosh, Sanjit Maitra, "Automatic detection and classification of diabetic retinopathy stages using CNN", International Conference on Signal Processing and Integrated Networks (SPIN), 2017

[6] Athira T R, Athira Sivadas, Aleena George, "Automatic detection of Diabetic Retinopathy using R-CNN", International Research Journal of Engineering and Technology (IRJET), Volume 6, Issue 5, May 2019

[7] Pranay Liya, Vaibhavi Shirodkar, Aashish Kapadia, "Detection of Diabetic Retinopathy using Convolutional Neural Network", Volume 6, Issue 4, Apr 2019

[8] Harry Pratt, Frans Coenen, Deborah M Broadbent, "Convolutional Neural Networks for Diabetic Retinopathy", ScienceDirect,

[9] Yung-Hui Li, Nai-Ning Yeh, "Computer-Assisted Diagnosis for Diabetic Retinopathy Based on Fundus Images Using Deep Convolutional Neural Network", Hindawi, Volume 2019

[10] Yashpal Shakti Kanungo, Bhargav Srinivasan, Dr. Savita Choudhary, "Detecting Diabetic Retinopathy using Deep Learning", IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017

Design of Intruder Detection System using Convolutional Neural Network and Raspberry Pi

Yadnesh Mohan Zagade, Mrs. Shiwani Gupta
Computer Enginnering
Thakur College of Engineering and Technology
Mumbai,India
zagade.yadnesh9@gmail.com

Abstract— **Security is an important part of our day to day life. It can be of your bank account, your home, your children etc. but the main thing is to check if your belongings are in safe zone. It is important to take prevention so that such crimes are avoided, rather than solving them after they take place. Thus, we propose a system using Raspberry Pi which shall be trained using Convolutional Neural Network (CNN) and deliver output using Rpi camera module and Bluetooth Speaker. Several researchers have worked in the area with respect to object detection and text to speech conversion, but many of them fail in their ability to identify the object. Thus, the proposed system detects unauthorized person, unwanted object, identifies it and speaks its class via Bluetooth speaker. The training shall be done using online coco dataset. The trained model shall be deployed onto Raspberry pi with camera module connected. Training shall be done with CNN the identified object class should be read out on Bluetooth speaker. The System is intended to be deployed for applications that require security with respect to the access into specific premises during day as well as night.**

Keywords: Neural Networks, Text to Speech Conversion, Image Processing, Features.

I. INTRODUCTION

J.

The field of Artificial Intelligence is essentially when machines can do tasks that typically require human intelligence. It encompasses Machine Learning, where machines can learn by experience and acquire skills without human involvement. Deep learning is a subset of Machine Learning where Artificial Neural Networks, inspired by the human brain, learn from large amounts of data. Similarly, to how we learn from experience, the deep learning algorithm performs a task repeatedly each time tweaking it a little to improve the outcome. We refer to 'deep learning' because the neural networks have various (deep) layers that enable learning. Just about any problem that requires "thought" to figure out is a problem deep learning can learn to solve.

The amount of data we generate every day is staggering currently estimated at 2.6 quintillion bytes and it's the resource that makes deep learning possible. Since deep-learning algorithms require a ton of data to learn from, this increase in data creation is one reason that deep learning capabilities have grown in recent years. In addition to more data creation, deep learning algorithms benefit from the stronger computing power that's available today as well as the proliferation of Artificial Intelligence (AI) as a Service. Organizations access to artificial intelligence technology and specifically the AI algorithms required for Deep Learning without a large initial investment. Deep Learning allows machines to solve complex problems even when using a data set that is very diverse, unstructured and inter-connected. The deeper learning algorithms learn, the better they perform.

Deep Learning architectures such as deep neural networks, deep belief networks and recurrent neural networks have been applied to fields including Computer Vision, Speech Recognition, Natural Language Processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases superior to human experts. Deep learning models are vaguely inspired by information processing and communication patterns in biological nervous systems yet have various differences from the structural and functional properties of biological brains (especially human brains), which make them incompatible with neuroscience evidences. Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised. Deep learning models tend to perform well with amount of data whereas old machine learning models stops improving after a saturation point.

II. Literature Review

K. Laubhan, et. al. [1], presented a system that consists of five ultrasonic sensors which continuously trigger serially and are positioned as left, right and front. The ultrasonic sensors are detecting objects and compares with the minimum value. If the value received is greater than threshold value then the user gets to know about the position of the obstacle via the Bluetooth headset. It successfully contributes in elimination of the walking sticks and detects if any object is present in surrounding of the blind person using the ultrasonic sensors positioned at five different directions hence the direction of object is also known. It detects position of the object, But it cant identify object.

Zhong-Qiu Zhao[2], presented a system on CNN which helps to determine where objects are located in a given image and which category each object belongs to (object classification). In this object detection models can be mainly divided into three stages: informative region selection, feature extraction and classification. They have deeper architectures with the capacity to learn more complex features than the normal ones. Also, this training algorithms allow to learn informative object representations without designing features manually.

Rithika.H1 , B. Nithya santhoshi[3] presents the system which basically can be used by people who do not know English and want it to be translated to their native language. The new thing of this research work is the speech output which is available in 53 different languages translated from English. This paper is based on a prototype which helps user to hear the contents of the text images in the desired language. This is done with Raspberry Pi and a camera module by using the concepts of Tesseract optical character recognition (OCR) engine, Google Speech API [application program interface] which is the Text to speech engine and the Microsoft translator.

Yeong-Hwa Chang[4] have used open source robotic middleware, it is applied to identify objects with a Raspberry Pi based mobile robot. Faster R-CNN algorithm very popular dataset for building and evaluating algorithms for image classification, object detection, and segmentation. COCO is a large-scale object detection, segmentation, and captioning dataset. The advantages about Microsoft COCO dataset provided for object segmentation, Recognition in context, multiple objects per image, more than 300000 images, more than 2 million instances and so on.

P. Patil, et. al. [5], presents a system that is easily wearable, user-friendly, and portable and can be used in their day to day life. This system consists of the belt along with modernized cane which duplicates the real sensation provided by the traditional white cane. It contains Raspberry pi 3, Ultrasonic sensors, PIR sensors, and Lithium battery. The ultrasonic sensor is used for distance measurement. It

the ultrasonic sound which is not audible to humans but can get reflected by obstacles and received by the sensor.

Muthukrishnan Ramprasath [6], presents system for Image Classification using Convolutional Neural Networks. This consist of uses deep learning algorithm to achieve the expected results in the area like computer visions. This system presents Convolutional Neural Network (CNN), a machine learning algorithm being used for automatic classification the images. By training the images using CNN network we obtain the 98% accuracy result in the experimental part it shows that this model achieves the high accuracy in classification of images.

Tiagrajah V. Janagiraman [7] presents research on “Traffic Light Detection Using Tensorflow Object Detection Framework”. They presents research on object detection for auto driving cars, where they have used CNN for object detection along with flexibility of Tensorflow Object Detection framework to solve real time problems. Its working on combination of CNN, faster RCNN with Tensorflow. They are working on giving more safe automated driving vehicle system. So they have presented automated driving system which will detect color of signal and will Start, Stop cars accordingly. This feature is not there in other driverless cars.

K.S.Shilpashree1 , Lokesha.H2 [8] presents paper about Implementation of Image Processing on Raspberry Pi. input is an image, like a photograph or video frame, the output of an image processing may be either an image or a video frame or a set of characteristics or parameters related to the image. Image processing concepts are implemented in Raspberry pi in the application of MAV. The system provides Debian Linux operating system Raspbian image for download. Python is used as main programming language for raspberry pi.

S. Venkateswarlu1 [9] in this paper has presented a robust approach for text extraction and converting it to speech Optical character Recognition (OCR) is a process that converts scanned or printed text images, handwritten text into editable text for further processing. Testing of device was done on raspberry pi platform. Picture is taken and processed by Raspy to hear the spoken words of the text through the earphone or speaker plugged into Raspy through its audio jack.

Yu Han LIU [10] presents recent studies about providing a technical solution for image recognition, by applying a algorithm called Convolutional Neural Network (CNN) which is inspired by animal visual system. Being highly-hierarchical networks, This paper talks about the basic understanding of CNN, including its structure, different layers, and working process. Feature extraction and redundancy reduction occur. The simple features gradually gather in an efficient way. Afterwards, all the features are combined partially and the resulting features each count for a part of configuration of the labelled class. Eventually, these top features are sent into the fully-connected layer and we get estimate of the classification.

ANALYSIS

Paper Title	Obstacle Detection	Object Identification	Wireless Screen Visual	Audio Output Wirelessly
A wearable portable electronic travel aid for blind[1]	Yes	No	No	No
Object detection using deep learning: A Review [2]	Yes	Yes	Yes	No
Image text to speech conversion in the desired language by translating with Raspberry Pi[3]	Yes	No	No	Yes
Deep learning for object identification in ROS-based mobile robots[4]	Yes	Yes	No	No
Environment Sniffing Smart Portable Assistive Device for Visually Impaired Individuals[5]	Yes	No	No	No
Image Classification using Convolutional Neural Networks[6]	Yes	Yes	No	No
Implementation of Image Processing on Raspberry Pi[8]	No	Yes	Yes	No
Text to Speech Conversion[9]	No	Yes	Yes	Yes
Feature Extraction and Image Recognition with Convolutional Neural Networks[10]	Yes	Yes	Yes	No

III.DESIGN OF PROPOSED SYSTEM.

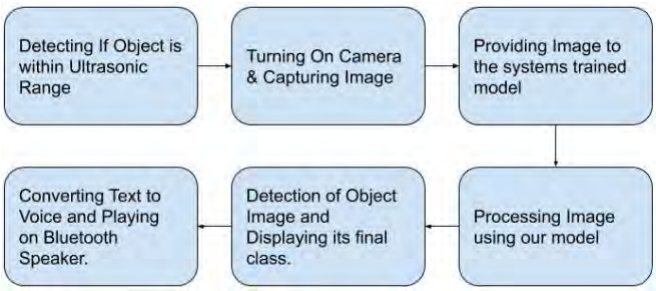


Figure 1: Block Diagram of the Proposed System.

If any intruder intersects within threshold range of ultrasonic system, then our camera will get turned on. It will capture all the things coming in its view and will send the images to our systems trained model that will process the image using trained model deployed onto Raspberry pi which is trained using CNN. Our system will identify the object and display its final Class on screen with accuracy. Accuracy is confidence of machine, that how much our system thinks that this is particular object.

Then text to speech will be done by doing extraction of text output, and it will speak out with the voice output on our Bluetooth speaker. The object identification will be done using CNN algo, by referring trained model on COCO dataset.

IV.METHODOLOGY

1. Data Collection

System totally works on neural networks that's on CNN, so our system need training dataset. In our system data will be in the form of images. Which we have to train to create a model, which we can use further when we are doing real time testing. We are going to use COCO dataset, for our object detection part of the system. It is a large-scale object detection, and captioning dataset. COCO has several Features. It has Object segmentation, Recognition in context, it also consist Super pixel stuff segmentation, total number of images used in this are around 330K (>200K labeled), have used approximately 1.5 million object instances, it consist of 80 object categories, 91 stuff categories, 5 captions per image. This dataset have different variants as well, which has different number of images for training. The dataset used in this project was an open source dataset.

2. Data Preprocessing

It is a process of transforming the raw, complex data into systematic understandable knowledge. It involves the process of finding out missing and redundant data in the dataset. On the basis of that the dataset is taken.

3. Building Model

Neural Networks is one of the most popular machine learning algorithms at present. It has been decisively proven over time that neural networks outperform other algorithms in accuracy and speed. With various variants like CNN (Convolutional Neural Networks), RNN(Recurrent Neural Networks), Auto Encoders, Deep Learning etc. neural networks are slowly becoming for data scientists or machine learning practitioners what linear regression was one for statisticians. It is thus imperative to have a fundamental understanding of what a Neural Network is, how it is made up and what is its reach and limitations. This paper is an attempt to explain a neural network starting from its most basic building block a neuron, and later delving into its most popular variations like CNN, RNN etc.

Increases with an increase in the number of layers and thus stagnates the learning of a neural network at a certain level.

In convolution layer the input is matrix form of image. In this the filter having particular weight multiplies matrix of image, and then creates output However, a sigmoid function also suffers from a problem of vanishing gradients. Sigmoid function squashes it's input into a very small output range

[0,1] and has very steep gradients. Thus, there remain large regions of input space, where even a large change produces a very small change in the output. This is referred to as the problem of vanishing gradient. This problem

The Convolutional Neural Network (CNN) is a class of deep learning neural network. CNNs represent a huge breakthrough in image recognition. They're most commonly used to analyze visual imagery and are frequently working behind the scenes in image classification. They can be found at the core of everything from Facebook's photo tagging to self-driving cars. They're working hard behind the scenes in everything from healthcare to security. Image classification is the process of taking an input (like a picture) and outputting a class (like "cat") or a probability that the input is a particular class ("there's a 90% probability that this input is a cat").

A CNN convolves learned features with input data and uses 2D convolutional layers. This means that this type of network is ideal for processing 2D images. Compared to other image classification algorithms, CNNs actually use very little preprocessing. This means that they can learn the filters that have to be hand-made in other algorithms. CNNs can be used in tons of applications from image and video recognition, image classification, and recommender systems to natural language processing and medical image analysis. It is important to select proper features to get desired output. Our machine learning algorithm is made for detecting different objects, in simple words for classifying the object it got in input, and giving its voice output. Machine learning fits mathematical notations to the data in order to derive some insights. CNNs have an input layer, and output layer, and hidden layers. The hidden layers usually consist of convolutional layers, ReLU layers, pooling layers, and fully connected layers. Convolutional layers apply a convolution operation to the input. This passes the information on to the next layer. Pooling combines the outputs of clusters of neurons into a single neuron in the next layer. Fully connected layers connect every neuron in one layer to every neuron in the next layer. CNN works by extracting features from images. This eliminates the need for manual feature extraction. The features are not trained! They're learned while the network trains on a set of images. This makes deep learning models extremely accurate for computer vision tasks. CNNs learn feature detection through tens or hundreds of hidden layers. Each layer increases the complexity of the learned features.

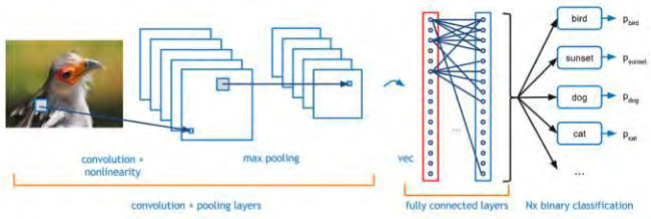


Figure 2: CNN Working

The depth of the filter has to be the same as the depth of the input, so if we were looking at a color image, the depth would be 3. That makes the dimensions of this filter 5x5x3. In each position, the filter multiplies the values in the filter with the original values in the pixel. This is element wise multiplication. The multiplications are summed up, creating a single number. If you started at the top left corner of your bubble wrap, this number is representative of the top left corner. Now you move your filter to the next position and repeat the process all around the bubble wrap. The array you end up with is called a feature map or an activation map! You can use more than one filter, which will do a better job of preserving spatial relationships.

5. Hardware requirements

Method is very necessary to get desired output, but along with software it also depends on hardware part of it. The hardware need was such that it should be compatible with CNN. Our coding for the same will be done in python, so the hardware is selected accordingly. Along with that we also used many different components and sensors compatible with above. The main aim of this paper is to propose system that helps the authorized person to get alert when something wrong is happens in his absence, or without his acknowledgement. So we will use method to convert text form of output into voice form. For that purpose we will be using processor compatible with Bluetooth, so that we can connect it to Bluetooth speaker. So our main hardware components will be Raspberry Pi 3b,RPi camera Module, Bluetooth Speaker, Power cord, Ultrasonic sensor.

V. CONCLUSION

Proposed system can be used for improving security of the Important places around us, such as home, office, building premises. This will help us to decrease crime rates. Above system will provide us hardware structure which can be installed easily at any required premises.

In future this system can be optimized to add few more features like monitoring and output through internet that is you can hear voice output from wherever you are. We can also put some notification to the user if something is identified using our system.

References

[1]K. Laubhan, et. al. “A wearable portable electronic travel aid for blind” DOI: 10.1109/ICEEOT.2016.7755039 Conference: 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)

[2] Zhong-Qiu Zhao “Object detection using deep learning: A Review” Published in IEEE Transactions on Neural 2018, DOI:10.1109/TNNLS.2018.2876865

[3] Rithika.H1 , B. Nithya santhoshi “Image text to speech conversion in the desired language by translating with Raspberry Pi” 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)

[4] Yeong-Hwa Chang “Deep learning for object identification in ROS-based mobile robots” 2018 IEEE International Conference on Applied System Invention (ICASI)

[5] P. Patil and A. Sonawane, “Environment Sniffing Smart Portable Assistive Device for Visually Impaired Individuals”, International Conference on Trends in Electronics and Informatics ICEI, 2017, pp. 317-321. https://bookauthority.org/books/best-rapsberry-pi-books 17

[6] Muthukrishnan Ramprasath, M.Vijay Anand “ Image Classification using Convolutional Neural Networks” ,International Journal of Pure and Applied Mathematics Volume 119 No. 17 2018, 1307-1319 ISSN: 1314-3395 (on-line version)

[7] Tiagrajah V. Janagiraman “Traffic Light Detection Using Tensorflow Object Detection Framework”, 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)

[8] K.S.Shilpashree1 , Lokesha.H2 , Hadimani Shivkumar “Implementation of Image Processing on Raspberry Pi” International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 5, May 2015

[9] S. Venkateswarlu1 *, D. B. K. Kamesh1 , J. K. R. Sastry2 and Radhika Rani “Text to Speech Conversion” Indian Journal of Science and Technology, Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/102967, October 2016

[10] Yu Han LIU “Feature Extraction and Image Recognition with Convolutional Neural Networks” First International Conference on Advanced Algorithms and Control Engineering IOP Publishing IOP Conf. Series: Journal of Physics: Conf. Series 1087 (2018) 062032 doi :10.1088/1742-6596/1087/6/062032

Image Classification Of Fasion-Mnist Dataset Using Cnn

Dhiiraj Yadav, Mrs Veena Kulkarni
Department of Computer Engineering
Thakur College of Engineering & Technology
Mumbai, India
erdhiraj1992@gmail.com, veena.kulkarni@thakureducation.org

Abstract— **Image classification is used in several applications, ranging from recognizing life-threatening illnesses in medical scans to detecting hotdogs in selfies. For humans, identifying numbers or items in a picture is extremely simple, but how do you train the machine to recognize these different things in images? Convolutional Neural Networks (CNN) can solve this problem. In this report, a convolutional neural network has been trained by identifying pictures in MNIST handwritten digital database to predict exactly what the numbers in the picture are. Obviously, human beings can perceive that there is a hierarchy or conceptual structure in the image, but the machine does not, for example the trained neural network is inconvenient to deal with special changes in a position of numbers in digital pictures. Exactly put, no matter what the environment of the image (image background) is, it is unchallenging for human beings to judge whether there is such a figure in the image and it is unnecessary to repeat the learning training. This dataset is publicly available for the research community to further advance handwritten digit recognition algorithms. Researchers at Zalando (the e-commerce company) have developed a new image classification dataset called Fashion MNIST in hopes of replacing MNIST. This new dataset contains images of various articles of clothing and accessories — such as shirts, bags, shoes, and other fashion items. The Fashion MNIST training set contains 55,000 examples, and the test set contains 10,000examples.**

Each example is a 28x28 grayscale image (just like the images in the original MNIST), associated with a label from 10 classes (t-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots).

Keywords— **Machine learning, CNN, Fashion, Image Annotation, E-ecommerce ,Image data, clothing, multiclass classification, shopping, object identification**

I. INTRODUCTION

Recently, digitization of handwritten documents has become significantly important to protect and store data more efficiently. The growth of digitized handwritten documents highlights new types of challenges and problems which lead to development of many automated and computerized analysis systems. Images are a special class of data that exhibit some interesting properties in respect to their

Structure. For example do the dimensions of an image (i.e. the pixel) exhibit a spatial relationship to each other Images are a special class of data that exhibit some interesting properties in respect to their structure. For example do the dimensions of an image (i.e. the pixel) exhibit a spatial relationship to each other?

The general idea is the following: if we want our model to generalize well, then we should design the learning process in such a way as to bias the model into learning such transformation-equivariant properties. One way to do this is via the design of the model itself, which for example was idea behind convolutional neural networks.

The MNIST data set is a classic handwritten digit recognition data set. This report shows how one can use conventional neural network to get a classification accuracy of 98%.

MNIST database is also widely used for training and testing in the field of machine learning.

The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems. The database is also widely used for training and testing in the field of machine learning. It was created by "re-mixing" the samples from NIST's original datasets. The creators felt that since NIST's training dataset was taken from American Census Bureau employees, while the testing dataset was taken from American high school students, it was not well-suited for machine learning experiments. Furthermore, the black and white images from NIST were normalized to fit into a 28x28 pixel bounding box and anti-aliased, which introduced grayscale levels.

II. BACKGROUND

The neocognitron was introduced by Kunihiko Fukushima in 1980. It was inspired by the above-mentioned work of Hubel and Wiesel. The neocognitron introduced the two basic types of layers in CNNs: convolutional layers, and downsampling layers. A convolutional layer

contains units whose receptive fields cover a patch of the previous layer. The weight vector (the set of adaptive parameters) of such a unit is often called a filter. Units can share filters. Downsampling layers contain units whose receptive fields cover patches of previous

In a variant of the neocognitron called the cresceptron, instead of using Fukushima's spatial averaging, J. Weng et al. introduced a method called max-pooling where a downsampling unit computes the maximum of the activations of the units in its patch. Max-pooling is often used in modern CNNs.

Several supervised and unsupervised learning algorithms have been proposed over the decades to train the weights of a neocognitron. Today, however, the CNN architecture is usually trained through backpropagation.

The neocognitron is the first CNN which requires units located at multiple network positions to have shared weights. Neocognitrons were adapted in 1988 to analyze time-varying signals

The name “convolutional neural network” indicates that the network employs a mathematical operation called convolution. Convolution is a specialized kind of linear operation. Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of a series of convolutional layers that convolve with a multiplication or other dot product. The activation function is commonly a RELU layer, and is subsequently followed by additional convolutions such as pooling layers, fully connected layers and normalization layers, referred to as hidden layers because their inputs and outputs are masked by the activation function and final convolution.

Though the layers are colloquially referred to as convolutions, this is only by convention. Mathematically, it is technically a sliding dot product or cross-correlation. This

Has significance for the indices in the matrix, in that it affects how weight is determined at a specific index point

When programming a CNN, the input is a tensor with shape (number of images) x (image width) x (image height) x (image depth). Then after passing through a convolutional layer, the image becomes abstracted to a feature map, with shape (number of images) x (feature map width) x (feature map height) x (feature map channels). A convolutional layer within a neural network should have the following attributes: Convolutional kernels defined by a width and height (hyper-parameters).

convolutional layers. Such a unit typically computes the average of the activations of the units in its patch. This downsampling helps to correctly classify objects in visual scenes even when the objects are shifted.

The number of input channels and output channels (hyper- parameter). The depth of the Convolution filter (the input channels) must be equal to the number channels (depth) of the input feature map.

Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus. Each convolutional neuron processes data only for its receptive field. Although fully connected feed forward neural networks can be used to learn features as well as classify data, it is not practical to apply this architecture to images. A very high number of neurons would be necessary, even in a shallow (opposite of deep) architecture, due to the very large input sizes associated with images, where each pixel is a relevant variable. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10,000 weights for

Each neuron in the second layer. The convolution operation brings a solution to this problem as it reduces the number of free parameters, allowing the network to be deeper with fewer parameters. For instance, regardless of image size, tiling regions of size 5 x 5, each with the same shared weights, requires only 25 learnable parameters. In this way, it resolves the vanishing or exploding gradients problem in training traditional multi-layer neural networks with many layers by using back propagation Convolutional networks may include local or global pooling layers to streamline the underlying computation. Pooling layers reduce the dimensions of the data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, typically 2 x 2. Global pooling acts on all the neurons of the convolutional layer. In addition, pooling may compute a max or an average. Max pooling uses the maximum value from each of a cluster of neurons at the prior layer. Average pooling uses the average value from each of a cluster of neurons at the prior layer.

III. MOTIVATION

Recently, digitization of handwritten documents has become significantly important to protect and store data more efficiently. The growth of digitized handwritten documents highlights new types of challenges and problems which lead to development of many automated and computerized analysis systems. Images are a

special class of data that exhibit some interesting properties in respect to their structure. For example do the dimensions of an image (i.e. the pixel) exhibit a spatial relationship to each other Images are a special class of data that exhibit some interesting properties in respect to their structure. For example do the dimensions of an image (i.e. the pixel) exhibit a spatial relationship to each other? The general idea is the following: if we want our model to generalize well, then we should design the learning process in such a way as to bias the model into learning such transformation-equivariant properties. One way to do this is via the design of the model itself, which for example was idea behind convolutional neural networks.

The MNIST data set is a classic handwritten digit recognition data set. This report shows how one can use conventional neural network to get a classification accuracy of 98%.

MNIST database is also widely used for training and testing in the field of machine learning.

IV. CNN

CNNs are regularized versions of multilayer perceptron's. Multilayer perceptron's usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to over fitting data. Typical ways of regularization include adding some form of Magnitude measurement of weights to the loss function. CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are on the lower extreme.

Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand- engineered. This independence from prior knowledge and human effort in feature design is a major advantage

When programming a CNN, the input is a tensor with shape (number of images) x (image width) x (image height) x (image depth). Then after passing through a convolutional layer, the image becomes abstracted to a feature map, with shape (number of images) x (feature map width) x (feature map height) x (feature map channels). A convolutional layer within a neural network should have the following attributes:

- Convolutional kernels defined by a width and height (hyper-parameters).
- The number of input channels and output channels

(hyper-parameter).

- The depth of the Convolution filter (the input channels) must be equal to the number channels (depth) of the input feature map.

Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus.Each convolutional neuron processes data only for its receptive field. Although fully connected feed forward neural networks can be used to learn features as well as classify data, it is not practical to apply this

Image recognition with CNNs

A system to recognize hand-written ZIP Code number involved convolutions in which the kernel coefficients had been laboriously hand designed.Yann LeCun et al. (1989) used back-propagation to learn the convolution kernel coefficients directly from images of hand-written numbers. Learning was thus fully automatic, performed better than manual coefficient design, and was suited to a broader range of image recognition problems and image types.

This approach became a foundation of modern computer vision. LeNet-5, a pioneering 7-level convolutional network by LeCun et al. in 1998, that classifies digits, was applied by several banks to recognize hand-written numbers on checks (British English: cheques) digitized in 32x32 pixel images. The ability to process higher resolution images requires larger and more layers of convolutional neural networks, so this technique is constrained by the availability of

computing resources.

Shift-invariant neural network

Similarly, a shift invariant neural network was proposed by W. Zhang et al. for image character recognition in 1988. The architecture and training algorithm were modified in

1991and applied for medical image processingand automatic detection of breast

cancer in mammograms .A different convolution-based design was proposed in 1988 for application to decomposition of one-dimensional electromyography convolved signals via de-convolution. This design was modified in 1989 to other de-convolution- based designs.

Neural abstraction pyramid

for each neuron in the second layer. The convolution operation brings a solution to this problem as it reduces the number of free parameters, allowing the network to be deeper with fewer parameters For instance, regardless of image size, tiling regions of size 5 x 5, each with the same shared weights, requires only 25 learnable parameters. In this way, it resolves the vanishing or exploding gradients problem in training traditional multi-layer neural networks with many layers by using back propagation architecture to images. A very high number of neurons would be necessary, even in a shallow (opposite of deep) architecture, due to the very large input sizes associated with images, where each pixel is a relevant variable. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10,000 weights

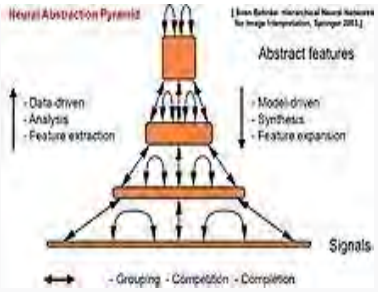


Fig: Neural Abstraction Pyramid

The feed-forward architecture of convolutional neural networks was extended in the neural abstraction pyramid by lateral and feedback connections. The resulting recurrent convolutional network allows for the flexible incorporation of contextual information to iteratively resolve local ambiguities. In contrast to previous models, image-like outputs at the highest resolution were generated, e.g., for semantic segmentation, image reconstruction, and object localization tasks.

GPU implementations

Although CNNs were invented in the 1980s, their breakthrough in the 2000s required fast implementations on graphics processing units (GPUs). In 2004, it was shown by K. S. Oh and K. Jung that standard neural networks can be greatly accelerated on GPUs. Their implementation was 20 times faster than an equivalent implementation on CPU. In 2005, another paper also emphasised the value of GPGPU for machine learning.

The first GPU-implementation of a CNN was described in 2006 by K. Chellapilla et al. Their implementation was 4 times faster than an equivalent implementation on CPU. Subsequent work also used GPUs, initially for other types of neural networks (different from CNNs), especially unsupervised neural networks.

In 2010, Dan Ciresan et al. at IDSIA showed that even deep standard neural networks with many layers can be quickly trained on GPU by supervised learning through the old method known as backpropagation. Their network outperformed previous machine learning methods on the MNIST handwritten digits benchmark. In 2011, they extended this GPU approach to CNNs, achieving an acceleration factor of 60, with impressive results In 2011, they used such CNNs on GPU to win an image recognition contest where they achieved superhuman performance for the first time Between May 15, 2011 and September 30, 2012, their CNNs won no less than four image competitions. In 2012, they also significantly improved on the best performance in the literature for multiple image databases, including the MNIST database, the NORB database, the HWDB1.0 dataset (Chinese characters) and the CIFAR10 dataset (dataset of 60000 32x32 labeled RGB images).

Subsequently, a similar GPU-based CNN by Alex Krizhevsky et al. won the ImageNet Large Scale Visual Recognition Challenge 2012. A very deep CNN with over 100 layers by Microsoft won the ImageNet 2015 contest.

V. APPLICATION

First, we need prepare out dataset. In this example, it will be the MNIST dataset. We need One-hot encode the labels of MNIST dataset. Each file has 1000 training examples. Each training example is of size 28x28 pixels. The pixels are stored as unsigned chars (1 byte) and take values from 0 to 255. The first 28x28 bytes of the file correspond to the first training example, the next 28x28 bytes correspond to the next example and so on. The example below loads the MNIST dataset using the Keras API and creates a plot of the first nine images in the training dataset.

According to the model given above, the final accuracy is stable at more than 90%. In fact, for many tasks, the model is prone to over-training. If over-training occurs during training, the error caused by training will decrease with the increase of training times.

The more data, the more data in the training set, the smaller the impact of training error and test error, and ultimately the accuracy can be improved. We can find the larger training set can develop the performance of LeNet5. After

improving this problem, our accuracy can reach 94%. If we increase the number of training, this accuracy will be improved to some extent.

This report is a completely new study for me. I chose the most famous handwritten digit recognition in the CNN model, which is also the earliest prototype of the convolution model. The datasets and reference papers were all from LeCun et al. This model is also the most classic model for image information processing. It also encountered some problems during the construction of LeNet-5 neural network. But those high-level interfaces for building neural networks in Python gave me a lot of convenience and came up with the training results mentioned in the references. But no doubt there is still much room for improvement in my identification of neural networks.

The digits patterns used in this training are all black and white images, so this model may be difficult to deal with common color patterns. The training set of color images may lead to an increase in the degree of ambiguity in the image, which in turn affects the accuracy of the prediction. Moreover, in this paper, I only studied the pictures in the MNIST data set. In actual problems, the specification and definition of the pictures are undoubtedly important factors affecting the accuracy of the model. Therefore, how to preprocess is very important, but in order to simplify the research process, I omitted this step. And in real life, most of the pictures we touch are very complicated and colorful

References

[1] [1] Gangaputra, Sachin. "Handwritten digit database". Retrieved 17 August 2013.

[2]

[3] [2] Qiao, Yu (2007). "THE MNIST DATABASE of handwritten digits". Retrieved 18 August 2013.

[4] LeCun, Yann; Corinna Cortes; Christopher J.C. Burges. "MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges". Retrieved 17 August 2013.

[5] Wan, Li; Matthew Zeiler; Sixin Zhang; Yann LeCun; Rob Fergus (2013). Regularization of Neural Network using DropConnect. International Conference on Machine Learning (ICML).

[6] [5] MNIST classifier, GitHub. "Classify MNIST digits using Convolutional Neural Networks". Retrieved 3 August 2018.

[7] [6] Ranzato, Marc'Aurelio; Christopher Poultney; Sumit Chopra; Yann LeCun (2006). "Efficient Learning of Sparse Representations with an Energy-Based Model" (PDF). Advances in Neural Information Processing Systems. 19: 1137–1144. Retrieved 20 September 2013.

[8] [7] Romanuke, Vadim. "Parallel Computing Center (Khmelnitskiy, Ukraine) gives a single convolutional neural network performing on MNIST at 0.27 percent error rate". Retrieved 24 November 2016.

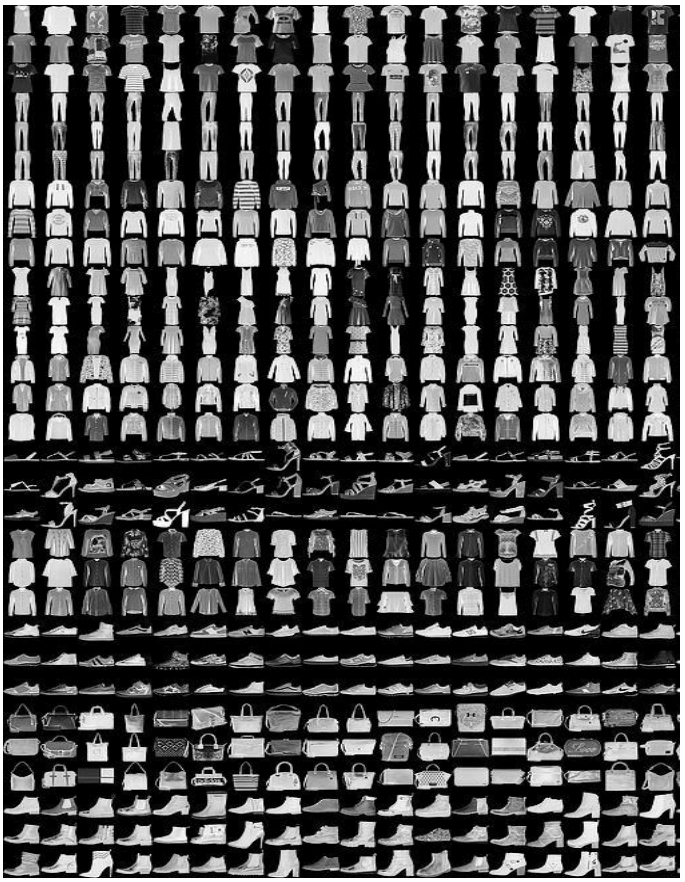
[9] [8] Patrice Y. Simard; Dave Steinkraus; John C. Platt (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. Proceedings of the Seventh International Conference on Document Analysis and Recognition. 1. IEEE. p. 958. doi:10.1109/ICDAR.2003.1227801. ISBN 978-0-7695-1960-9.

[10] [9] DeCoste and Scholkopf, MLJ 2002

[11] [10] Kégl, Balázs; Róbert Busa-Fekete (2009). "Boosting products of base classifiers" (PDF). Proceedings of the 26th Annual International Conference on Machine Learning: 497–504. Retrieved 27 August 2013.

[12] [11] LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). Gradient-based learning

[13] applied to document recognition. Proceedings of the IEEE, 86(11):2278-2324.



An Efficient Privacy Preserving Technique And System For Data Publishing On Social Media

Snehal Tukaram Patil, Dr. Aand Khandare

Thakur College of Engineering & Technology ,

Kandivali(E),Mumbai ,Maharashtra ,India

snehal321.patil@gmail.com, anand.khandare1983@gmail.com

Abstract—**In the era of technology and network, big data and huge amount of data is produced over the world-wide. The use and sharing of enormous amounts of data has become possible. Once the information and data accumulated, it will become the wealth of information. The existing traditional data processing techniques and algorithms directly operated on the initial data set, which is able to cause the leakage of privacy data. At the same time, huge amounts of information implicate the sensitive knowledge that their disclosure cannot be ignored to the competitiveness of enterprise. So as to beat these problems, Privacy Preserving Data Mining and Data Publishing techniques are developed. Traditional PPDP techniques suffer from differing types of attacks and loss of data.**

In this paper new method is proposed that improves the performance of the present & traditional PPDM techniques. The comparison of existing techniques and proposed technique is additionally discussed in the paper and results shows the time taken by the proposed system is very less than the existing techniques. The proposed method provide less information loss and more privacy.

Keywords— Information Filtering, Social Computing, Social Network, Topic Identification, Topic Ranking, leakage of privacy, key based encryption.

I. INTRODUCTION

Due to World Wide Web, there is vast amount of information available over the network. This information could be accessed through data mining. Data mining is concerned with the extraction of nontrivial, novel and potentially useful knowledge from large databases. In order to extract the knowledge various data mining techniques are used as per the application domains like health care, Cyber security, banking, e-commerce etc. [1]. These domains contain confidential data which should not be disclosed to all users which lead to the development of privacy preserving techniques in data mining [2]. PPDM is a new research in data mining where data mining functionalities are analysed with respect to privacy in data.

One of the central problems of managing privacy within the Internet lies in the simultaneous management of both public and personal data. So as to receive services like recommendations, users often willing release some data about themselves, like their movie watching history. However users even have other data they consider private, such as income level, political affiliation, or medical conditions.

In this work, we focus on a technique in which a user can release her public data, but is able to prevent against inference attacks. With the currently growing interest in the Semantic Web, it is reasonable to expect that increasingly more metadata describing domain information about resources on the Web will become available. The idea presented here is to enrich the search process for hypermedia applications with information extracted from the semantic model of the application domain. One of the major problems of traditional search engines is that they typically are based in keyword processing.

In this paper, the key based encryption algorithm is proposed which protect user activity data, like the privacy leakage of user-specified private data is minimized. The system achieves better privacy protection and less leakage of private data.

II. LITERATURE SURVEY

Today many organizations publish their data. This information contains private data of individual, organization, industry, and person. Public information should not be released without preserving the privacy of each record. It should ensure that no individual information is identified from the anonymized data. Anonymization approaches has been utilized in numerous application domains. Social networks have always existed in society in varying forms. People are unaware of the privacy issues. The published data in the social network can be anonymized and do not infer to any single individual.

Large organizations in different sectors viz. medical, social networking, banking, and business, insurance collect, store and use personal information of their customers. This collected data are used for the research and analysis purposes [2]. To do the same, data mining techniques have been used for the analysis and research work. The collected data might contain person specific private information. Therefore, analyzing such collected person specific private data can disclose the private information of an individual. Protecting the private information of an individual becomes a prime research issue in privacy preserving data publishing. The user data is very important and need the strong privacy. By considering the factors there are many privacy models in the privacy preserving data publishing. Privacy models which gives different algorithms and different techniques for achieving privacy like Anonymization which contains the techniques such as generalization, bucketization, slicing. K-anonymity and L-diversity.

In this section we will be going to discuss about the above mention existing techniques, and the proposed key based privacy preserving hash function technique.

A. K-Anonymity

If the identifiable attributes of any database record are undistinguishable from at least other k-1 records, then the dataset is known as k-anonymous [7]. In k-anonymity

attributes are suppressed or general awaiting every line is the same through at slightest k-1 extra rows. At this point the record is supposed to be k-anonymous. It can separate access details and non-access details of Data set. It'll include slicing approach of Column separation method.

B. L-diversity

The L- diversity can be described as Say if you have a group of k different records that all share a particular quasi-identifier. That's good, therein an attacker person cannot identify the individual based on the quasi-identifier. But if the worth they are fascinated by, (e.g. the individuals' medical diagnosis) is that the same for each value within the group. The distribution of target values within a bunch is referred to as l-diversity. At present, there exist two categories of l-diversity techniques: generalization and permutation-based. The current generalization method would partition the data into disjoint groups of transactions, specified each group contains sufficient records with l-distinct, well represented sensitive items. It preserves both record and attributes linkage.

The Table 1 shows the k anonymity =6 and L-diversity =4, it means only six rows be able to be position in one slice table view with L-diversity have apply on disease attribute. The four unique values has place in diversity attributes.

Table 1: K-anonymity and L-diversity

Name	Address	Zip code	Age	Disease
****	Nairobi	452***	[0-25]	cancer
****	Melbourne	145***	[26-50]	cancer
****	Sidney	365***	[26-50]	cancer
*****	Jakarta	356***	[51-75]	diabetes
*****	Mumbai	458***	[26-5]	angina
*****	Delhi	564***	[51-75]	cancer

C. SHA 256 Hash Function

Large organizations in numerous sectors viz. medical, social networking, The SHA (Secure Hash Algorithm) relations be a set of connected cryptographic hash function designed by the algorithm creates a hash value from any kind of data, such as a file, password, or in this case, a praise certificate number. This value is virtually unique to the input data, so even a small change in the data will bring about a totally extraordinary hash because of the torrential slide impact. Additionally, there is no down to earth approach to compute a specific information input that will bring about a greatly desired hash esteem and it is very difficult to utilize the hash an incentive to recuperate the first information. The most generally utilized capacity in the family is SHA-1 and it is utilized in a vast assortment of prevalent security applications and conventions. Things being what they are, the reason is it still safe to utilize SHA-1 to encode your Master card numbers? Since the odds of two charge card numbers having a similar hash esteem are so little, it's far-fetched you'll discover a circumstance where the hash of a decent card number matches the hash of a terrible card number, accordingly discounting the likelihood of a false positive. The Table 2 shows SHA-256 results of plain data.

Table 2: SHA 256 view

Name	Address	Zip code	Age	Symptoms	Disease
John	Nairobi	452****	[0-25]	398ed52a42b0fd22abb8252393a73e29	cancer
Ram	Goa	145****	[26-50]	bcf82b421a0beb03c7d682fada277afa	cancer
Jhon	Sidney	365****	[26-50]	c7e36c2e7d63e53228831b18a61a092f	cough
Khan	Jakarta	356****	[51-75]	e4b87007c431335a0f27b0cd1c1597bc	diabetes
RJ	Delhi	145****	[26-50]	bcf82b421a0beb03c7d682fada277afa	cancer
Shah	Pune	365****	[26-50]	c7e36c2e7d63e53228831b18a61a092f	cough

Similarly, the assault is an impact assault, not a pre-picture assault. As I beforehand said, an impact assault discovers two bits of information with a similar hash, however the aggressor can't pick what the hash will be and thusly can't break the instruments that utilization SHA-1 to check for changes in a hashed information. Then additionally, a pre-picture terribly empowers somebody to locate a terrible Visa number that causes a hash capacity to deliver a hash estimation of a substantial card number. Nonetheless, in light of the fact that you are utilizing a boycott, the aggressor can't exploit this, on the grounds that the examination procedure would locate the awful card number on the reject list. On the other hand there is less chance that you are as yet concerned, you could consider utilizing SHA-224, SHA-256, SHA-384 or SHA-512, here and there all things considered referring in indirect way to as SHA-2. This would require additional storage room, notwithstanding, in light of the fact that SHA-1 makes a hash value size of 160 bits and SHA-224, for example, makes an estimation of 224 bits. Likewise, the correlation procedure would be a little slower.

III. System Design

In the proposed research work to design and implement a system than work as classify and re-rank all kind of query events along with the present events further as news. The Google API will provide the third party interface for communicate with search engine to classify the all data using machine learning approach and re-rank with page rank and click through algorithms. We are collect the information from Google API, YouTube as well as twitter and rank all the news base on current user query

A. Project Overview

The proposed system consolidates and rank the foremost prevalent topics discussed in both journalism and social media during a selected period of time. To gain its goal, the system must undergo four main stages.

Pre-process and features extraction: In this module system extract all the information shared by user and apply similarity base weight calculation using private and public

training rules. For each attribute we will generate the specific weight according to training policies.
Classification using privacy framework: This module will classify all data as private or public according to current similarity weight, the data which is classified by classification algorithm as public or private those attributes which is shared by end user.
Re-Ranking: Here system choose to measure the pair wise ranking loss using a widely known metric, created two private and public vectors to different users using Kendall-T distance algorithm.
Analysis: System provide the comparative analysis using different parameters for accuracy time complexity using different graphs.

B. Development Methodology

The above figure 1 shows a proposed system architecture with execution process, in first phase user uploads the his own information on social media, it contains some private as well as some public information. The proposed work also carried out to identify the potential of private data and provide auto security using proposed given algorithms. Natural Language Processing (NLP) extract the features from user's shared data and apply the weight clustering using cosine similarity algorithm. Once clustering has done it will generate the three different clusters viz public, private and activity etc.

The second phase in system implementation of ranking algorithm according to their weights and apply the privacy policies on those private data attributes. In whole process system automatically maintain the privacy on vulnerable data which is shared by data owner.

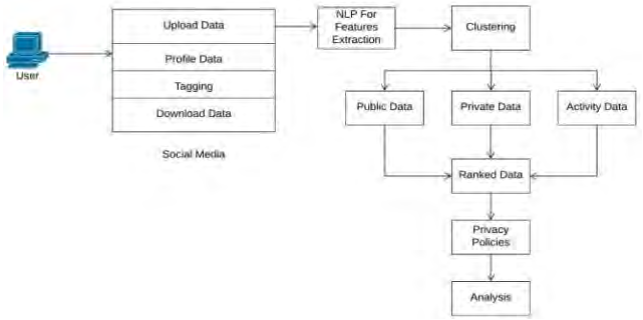


Figure 1: Proposed System Architecture

IV. PROPOSED TECHNIQUE

In a system that not only records but also encrypts the video, before storing it for later use. The idea is that while the officer is on duty, the camera will always be recording. The video, however, is broken up into smaller segments for efficiency. After a certain time period passes, the video that was recorded is then encrypted using an encryption key so as to automatically generate. Certain metadata, including the date as well as time of the video, the name of the video, the encryption key, and the name of the client is saved for later use. This procedure be repetitive used for as long as the officer is on duty. At the ending of their shift, the officer archives the videos to the backend server. The backend server will execute classification algorithms which receives the encrypted videos, decrypts them, and classifies them based on their content and the metadata that was previously saved. Classification criteria include whether the video was recorded inside a safe position, whether the video captured a crime and the type of crime recorded, and whether the video contains sensitive content. This classification is then used to perform additional encryption for access control. When an officer wishes to access the videos, he will log into the retrieval program, which will assign the appropriate decryption keys base on the request. If the officer has the appropriate keys, he will might be get back and decrypt the necessary videos. The below Table 3 shows the key base encrypted view.

Table 3: Key based encrypted view

Name	Address	Zip code	Age	Symptoms	Disease
John	Nairobi	452***	[0-25]	06F6620636F726F6E6172792	cancer
Ram	Goa	145***	[26-50]	646F6D697A6174696F6E20737	cancer
Jhon	Sidney	365***	[26-50]	656C20363030206D6720666	cough
Khan	Jakarta	356***	[51-75]	7374656E6F7369730A2020202045	diabetes
RJ	Delhi	145***	[26-50]	41726D2920696D706C616E	cancer
Shah	Pune	365***	[26-50]	E742053797374656D0A20	cough

In key base hash function it generates very much long encrypted text, which is related to privacy approaches like, but the main different in key base hash encryption is it generate probable cipher data up to 256 bytes.

A. Key base hash privacy algorithm

Input: Generate Key, data d
Output: Generated hash H according to given data
Step 1: Input data as d
Step 2: Generate secure random key from byte[].
Step 3: Apply Secure Hash from Hash family
Step 4: CurrentHash= SecureHash (d,key)
Step 5 : RetrunCurrentHash

Observations:

- It provides secure data broadcasting without any secure multiparty computation, so it reduces the execution cost.
- It doesn't support for secure data extraction from privacy view, it means there is no possibility for any kind attacks like collusion, bruited forces attack etc.
- It is applicable on structured as well as semi structured dataset respectively.

V. COMPARATIVE ANALYSIS

Algorithms Name	Proposed	Observations
K-Anonimity, L-Diversity, Anonimization, Generalization, Random Permutation, Generalization, Specialization, Slicing, Bucketization etc	Existing	Very high time complexity for multi-party computation, and generate secure view It work only data distribution time. Original data can be leaked by attacker , because its privacy provide sufficient Background Knowledge (BK) to attacker.
PrivRank base privacy framework using obfuscate	Base Paper	Cluster generation follow long process, and generate very high time complexity, According base system privacy parameter has extracted based on mutual information between released public and private data, it can be leaked sensitive inform during the data broadcasting. It is too much hard to apply on different kind sensitive as well as private parameters values like health care privacy data etc.
Key base hash function	Proposed (Key Contribution)	It provides secure data broadcasting without any secure multi party computation, so it reduces the execution cost. It doesn't support for secure data extraction from privacy view, it means there is no possibility for any kind attacks like collusion, bruited forces attack etc. It is applicable on structured as well as semi structured dataset respectively.

VI. RESULTS AND DISCUSSIONS

This section discuss the dataset used and result obtained from the modified algorithm. It also compares it with the existing algorithm. Result obtained from the algorithm have been checked on different types of the datasets. The dataset used in system text format dataset.

searchword	twitteruser_name	twitterdate	twitterpost	url	scorecs
rahul gandhi	wish4u52	Tue Apr 27 18:00:19 IST 2010	RiaRevealed BJP4India INCIndia RahulGandhi All nonsens...	null	0.3535533905932737
rahul gandhi	jay_ambadi	Wed Jan 28 12:30:01 IST 2009	And they say Rahul Gandhi is insignificant!	null	0.5345224838248487
rahul gandhi	INCTharoorian	Sun Jun 21 13:15:14 IST 2009	RT INCTharoorian: RahulGandhi Try imagining Modi being...	https://t.co/JQ75aHSeTFI	0.3244428428152503
rahul gandhi	INCTharoorian	Sun Jun 21 13:15:14 IST 2009	RahulGandhi Try imagining Modi being in the picture inste...	https://t.co/JQ75aHSeTFI	0.3429971702850764
rahul gandhi	dc79131542	Thu Nov 22 15:26:23 IST 2018	Technology Disaster Rahul Gandhi setting the voice re...	null	0.1666666666666666
rahul gandhi	AshishKG2014417	Sat Jul 28 21:57:37 IST 2018	Protesting against the Rafael deal Rahul Gandhi & Adhir ...	null	0.3429971702850764
rahul gandhi	manojmishraom	Thu Jun 17 14:49:45 IST 2010	RT MaheshJoshi_MJ: Apologizing is not new for Rahul Ga...	RahulRaFAL	0.3922232270276368
rahul gandhi	Ankush065021363	Fri Oct 26 08:36:43 IST 2018	Another contempt of SC by Rahul Gandhi & Ajay Shukla(...	null	0.4866642632622875
rahul gandhi	aashish2212	Thu Apr 24 19:04:11 IST 2014	ANI: Why dont u guys protest against the pollution,Declini...	null	0.1666666666666666
rahul gandhi	bkmshr	Fri Oct 03 00:58:32 IST 2014	serious_joker1 Sarita251487177 Rahul Gandhi attempted t...	null	0.3244428428152503
election maharashtra	IndiaToday	Mon Feb 02 12:51:54 IST 2009	The challenge looks particularly daunting in Jharkhand in ...	null	0.2828427124746119
election maharashtra	kuchpatabetado	Tue Aug 28 17:29:51 IST 2018	RT mdaleembasha: Next election BJP address gull	null	0.2499999999999994
election maharashtra	mdaleembasha	Mon Nov 18 15:50:52 IST 2013	Next election BJP address gull	null	0.2886751345948129
election maharashtra	vishalkodvani7	Sat Dec 20 09:27:09 IST 2014	thakur_shivangi Ek Seat Khaali Krvani Padegi... Ek Electi...	null	0.1507556722888818
election maharashtra	p_jakumar	Fri Jan 07 20:44:26 IST 2011	Shiv Sena BJP| Maharashtra: Internal revolt in Shiv Sena...	null	0.1507556722888818
election maharashtra	KEYESEN2000	Sat Jan 09 22:10:47 IST 2010	RT mugilkar_: Maharashtra: Internal revolt in Shiv Sena? ...	null	0.1714985851450882
election maharashtra	LokeshG118 19168	Mon Apr 22 22:17:27 IST 2019	RT ZeeNews: NCP-Shiv Sena-Congress govt will last five ...	null	0.1474419561549871
election maharashtra	Yogendr15313428	Mon Nov 28 13:49:18 IST 2016	RT ZeeNews: NCP-Shiv Sena-Congress govt will last five ...	null	0.1474419561549871

fig 1: Sample Dataset

This section discuss the dataset used and result obtained from the modified algorithm. We focus on ranking stage that considered as the fundamental commitment of this paper. New positioning calculation is delivered to rank comparative significant information in the wake of ordering stage. Notwithstanding, information recovery process turn out to be quicker, less demanding and more exact. The execution accomplished with 99 percent applicable outcomes in greatest time 60 ms and 1 percent just for immaterial outcomes. The proposed structure and positioning calculation can be additionally produced for later use in distinguishing more exact semantic data from interpersonal organizations in a brief timeframe. The point of the semantic web index has pulled in substantial interests both from industry and research with coming about assortment arrangements in various errands. There is no institutionalized structure that screens and fortify the advance in this field. In this paper, Four standard undertakings of semantic web crawler are talked about including creeping, ordering, positioning lastly recovering assignment. This survey paper demonstrates that utilizing PrivRank produces an altogether different positioned rundown of news themes, which may imply that depending just on high-recurrence news points gave by the media does not really give understanding into what clients are intrigued on or think about essential. Contemplating all outcomes underlines the point that MF alone is a substandard estimator of what clients find fascinating or think about imperative, and ought to thusly not be utilized as a part of along these lines. PrivRank, then again, turns out to be more equipped for playing out this, thus we infer that the data gave by PrivRank can demonstrate imperative in business based territories where the enthusiasm of clients is paramount.

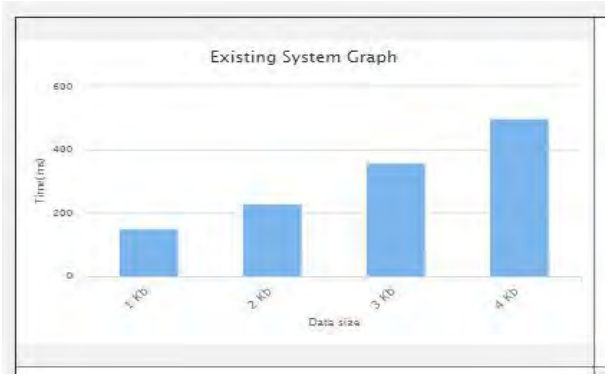


fig 2 (a)

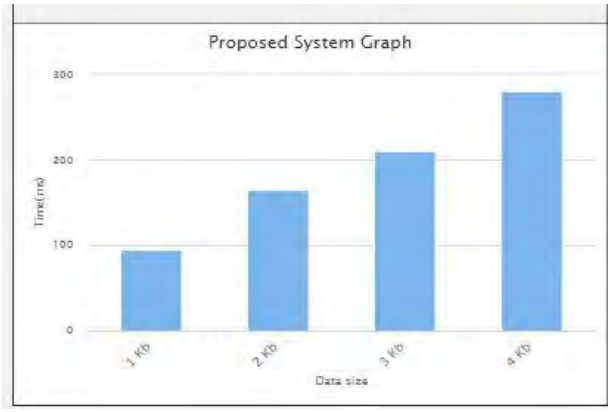


fig 2 (b)

fig 2: Execution time comparison between proposed vs. existing approach

Figure 2 shows the comparison between the existing technique or methods for privacy preserving and the proposed key based hash algorithm. From the Fig 2(a) & Fig 2(b) it is observed that the time require to compute the privacy function on data is reduced by the proposed technique.

As well the privacy gain has been better within the proposed method compared to the traditional method.

CONCULSION

Privacy-preserving distributed information/data mining could be a siblings to the research topic of privacy-preserving data publishing (PPDP). PPDDM assumes a scenario that more than one data holders want to collaboratively perform information processing on the union of their data without disclosing their sensitive information. PPDDM usually employs cryptographic solutions. Sharing of information has become part of the routine activity of the numerous individuals, companies, organizations, and government agencies. Privacy-preserving data publishing is a approach to data sharing, while preserving individuals' personal and protecting sensitive information. Information sharing has become a part of the routine activity of the many individuals, companies, organizations, and government agencies. Privacy-preserving data publishing is a promising approach to data sharing, while preserving individual personal data and protecting sensitive information.

System proposed an unsupervised technique Automatic privacy preservation on public additionally private data. The privacy function which is able to filter all user data and apply anonymity base privacy. Automatic Role Base Access Control to non-public parameters for user data. It also provide ranked data in line with to user's perspective for each new session.

In this paper, we assesses the traditional methods in the field. The goal is to convert the original data into some anonymous form to prevent from inferring its record owners' sensitive information. Most previous privacy-preserving techniques were proposed for data publishers, but individual record owners should also have the right and responsibility to protect their own private information.

References

[1] CH Venkata Lakshmi, Swapna Siddamsetti "An enhanced approach for privacy preserving data mining(PPDM)", *CETCSE*, 2017

[2] Magdalini Eirinaki, Michalis Vazirgiannis "Usage-based page rank for web personalization" *Proc. fifth IEEE international conference on Data mining (ICDM "05)*

[3] Samriti Gupta, Alka Jindal "Contrast of Link based Web Ranking Techniques" *IEEE Proc.* 2015.

[4] Chi Chen at. Al. proposed "An Efficient Privacy-Preserving Ranked Keyword Search Method" *IEEE* 2016.

[5] Jun fang,LeiGuo, "Calculation of weight of entities in ontologies by using usage based information" *IEEE Proc.of* 2011.

[6] M. P. Selvan, A.ChandraSekar "Priya Darshan Survey on web page ranking algorithms*" International Journal of Computer Applications,* 2012.

[7] Zhicheng Dou and Zhengbao Jiang, "Proposed Automatically Mining Facets for Queries from Their Search Results" *IEEE transaction and data engineering,* vol 28, 2017.

[8] Nissim Matatov, Lior Rokach and Oded Maimon "Privacy preserving data mining: A feature set partitioning approach" *IEEE transaction and data engineering,* 2010.

[9] Ashlesha Gupta, Ashutosh Dixit, AK Sharma "Relevant document crawling with usage pattern and domain profile based page ranking" *IEEE Proc*, 2013.

[10] Zhao, B., et al., "TEXplorer: keyword-based object search and exploration in multidimensional text databases", *Proc. of the 20th ACM Conference on Information and Knowledge Management (CIKM)*. ACM: Glasgow, Scotland, UK. p. 1709-1718, 2011

[11] Shital C Patil,RRkeole "Content and usage based ranking for enhancing search engine delivery" *issue 7th International journal of Science and Research (IJSR),* vol 3, 2014.

[12] Safaa I. Hajeer, Rasha M. Ismail, Nagwa L. Badr, M. F. Taiba "An Efficient Hybrid Usage-Based Ranking Model for Information Retrieval Systems and Web Search Engines", *IEEE Proc. of 6th International Conference on Information and Communication Systems (ICICS), 2015.*

[13] AzamFeyznia, mohsin Kahanti "A link analysis based ranking method for semantic web documents" *IEEE Proc.,* 2013.

[14] Samarati P, Sweeney L. Protecting "Privacy when Disclosing Information:k-Anonymity and its Enforcement through Generalization and Suppression", *IEEE Symp. on Security and Privacy,* 1998.

[15] G.K.Panda, A.Mitra, Ajay Prasad, Arjun Singh, Deepak Gour, "Applying l-Diversity in anonymizing collaborative Social network", *(IJCSIS) International journal of Computer Science and Information Security,* vol. 8, Issue 2, 2010.

[16] Chen Li, Houtan Shirani-Mehr, Xiaochun Yang "Protecting individual information against inference attacks in Data Publishing" *International Journal Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557{570, 2002.

[17] Mingqiang Xue, Panagiotis Papadimitriou, Chedy Ra¨ıssi, Panos Kalnis, and Hung Keng Pung " Distributed privacy preserving data collection" *In: Yu J.X., Kim M.H., Unland R.(eds) Database Systems for Advanced Applications. DASFAA,v*ol 6587,2011

[18] Amita Sharma and N. Badal, "Literature Survey of Privacy Preserving Data Publishing (PPDP) Techniques", *International Journal Of Engineering And Computer Science,* vol. 6, pp. 21258-21268, 5 May 2017.

[19] M. Raya, J.-P. Hubaux, "The security of vehicular ad hoc networks", *Proc. SASN*, pp. 11-21, Nov. 2005.

[20] X. Lin, X. Sun, X. Wang, C. Zhang, P. H. Ho, X. Shen, "TSVC: Timed efficient and secure vehicular communications with privacy preserving", *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4987-4998, Dec. 2008.

[21] X. Lin, X. Li, "Achieving efficient cooperative message authentication in vehicular ad hoc networks", *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3339-3348, Sep. 2013.

Distributed Data Mining

Deepali Joshi

Assistant Professor

Thakur College of Engineering & Technology

deepali88.dj@gmail.com

Abstract:

Human beings are surrounded by numerous devices and is almost impossible to complete work without these devices. In this process the devices perform various operations and utilize, create numerous data. Devices are present at numerous locations that is they are distributed. The data generated and utilized by these devices should be analyzed. All the data that is generated by the devices may not be useful, So Data mining technique to help to extract essential data. The data may be located at different locations and due to this Data mining should be done with respect to distributed devices. This process is known as Distributed Data Mining (DDM).

Keywords
Data Mining, Distributed Data Mining

I. INTRODUCTION

In the recent years there is a boom in data generation. Number of devices are utilized by users. The devices can be wired or wireless. The enhancement in data mining leads to distributed data mining (DDM) that will mine data from different sources available at different locations. The best suitable example is internet, sensor which is commonly used and spread across large geographical area. Distributed Data Mining (DDM) can be described as extraction of essential data, information, knowledge, patterns from large amount of distributed databases.

II. Architecture

2.1 Data Mining Architecture

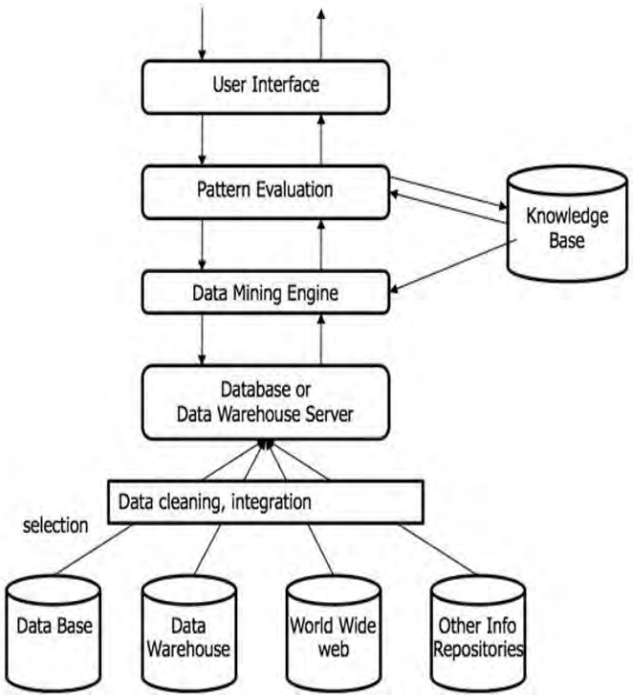


Fig 1: Data Mining Architecture

1. Database, data warehouse, www, other information repository

These are repositories where the data can be stored. The data which is required will be searched in repositories. Once the data is selected number of steps are to be completed. The data may have some missing values or may have incorrect data so Data cleaning is to be performed. The data may be gathered from different sources and will be known as Integration.

2. Database or data warehouse server

It is responsible for fetching essential data, based on user's data mining request.

3. Data Mining Engine

The data mining engine is responsible for performing data mining operations on the data. Depending upon the application appropriate data mining algorithm will be selected. It consists of set of functional modules for tasks like classification, prediction, etc.

4. Knowledge base

This is domain knowledge that guides the search or evaluate interestingness of resulting patterns. Such knowledge can be used to organize attributes into different levels (hierarchies).

5. Pattern Evaluation Module

It interacts with the data mining module so as to focus the search toward interesting patterns. It can be integrated with data mining module.

6. User Interface

It communicates between users and the data mining system, allowing the user to interact with system by specifying a data mining query or task. It allows the user to browse database and data warehouse schemas or data structures and visualize the patterns in different forms.

2.1 Distributed Data Mining Architecture

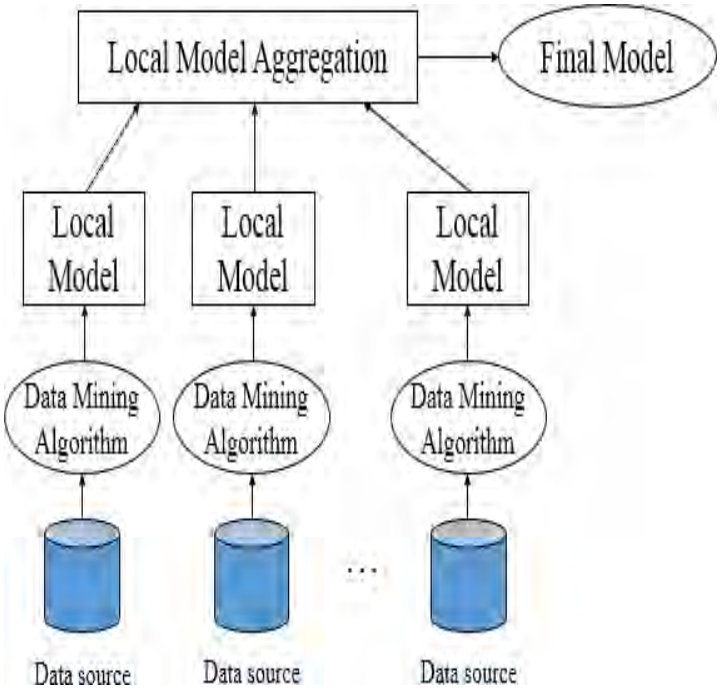


Fig 2: Distributed Data Mining Architecture

In distributed data mining, there are local data sources at every location. On this data source required data mining algorithms are applied. The result that is generated is aggregated together and then the final result is evaluated. The data may be downloaded to single site and then data mining operations may be performed. In order to perform the operations it is very essential to know how exactly the data is distributed. The data can be in different forms, so the first step is to identify it. Distributed data mining is executed in parallel manner. The same algorithm is applied to data that is available at different locations.

III. ISSUES IN DISTRIBUTED DATA MINING

3.1Replication of Data: Local databases are available at each location and similar data may be available.

3.2. Fragmentation of Data:

Local tables are present in each location. Tables are divided into rows and columns. Fragmentation can be horizontal or vertical.

3.1. Communication Cost:

In distributed environment, the devices which are situated in different locations have to communicate with each other for transferring data, command.

3.2. Integration of results:

The results of distributed data mining systems are collated from various sources and this operation may be difficult. The result is combination of outputs generated by different models.

IV. APPLICATIONS OF DISTRIBUTED DATA MINING

- Credit card fraud detection
- Intrusion detection
- Information retrieval from Internet
- Ad hoc sensor networks

V. CHALLENGES OF DISTRIBUTED DATA MINING

- Real-time distributed data mining
- Adaptive to changing environment, new data, new pattern

VI. CONCLUSION

Distributed Data Mining gathers data from various sources and then it is integrated into one component. So it can be said that the operations are carried out on different local machines simultaneously and then combined to generate a result. The data mining architecture is of centralized form where operations are carried out on local data whereas in case of distributed data mining, the operations are carried out on local data and then the patterns are evaluated by combining data from different sources. The main advantage is that even though data is available on different machines, it can be integrated.

References:

[1]. Rupali Chikhale 'Study of Distributed Data Mining Algorithm and Trends' IOSR e-ISSN: 2278-0661,p-ISSN: 2278-8727

[2]. Byung-Hoon 'Distributed Data Mining : Algorithm, Systems and Applications'

[3]. http://www.csee.umbc.edu/~hillol/PUBS/review.pdf

[4]. http://www.distributeddatamining.org/

[5]. http://www.dfki.de/~klusch/papers/eaai.pdf

[6]. http://link.springer.com/article/10.1007%2Fs10799- 012-0124-y#page-1

[7]. http://neuron-ai.tuke.sk/babik/19310438.pdf

[8]. http://dml.cs.byu.edu/~cgc/docs/atdm/Readings/Distributed DM.pdf

Power Of Visualization In Data Analysis

Anupriya Babbar
Assistant Professor
Thakur College of Engineering and Technology
Mumbai 400101 India
anupriyababbar.86@gmail.com

Abstract— **Data Visualization is a powerful tool, it turns monotonous information into a visual that our brains can actually relate and understand. Because our brains process visuals up to 60,000 times faster than text. Visualization provides deep insight about data. There are many powerful libraries like pyplot.Matplotlib, Seaborn, bokeh of python**

Which build interactive graphs, visualization involves producing images that communicate relationships among the represented data to viewers of the images. It helps in decision making and gives deep insight about the data and its relation with the other attributes.

Keywords— Data Visualization, Data Analytics

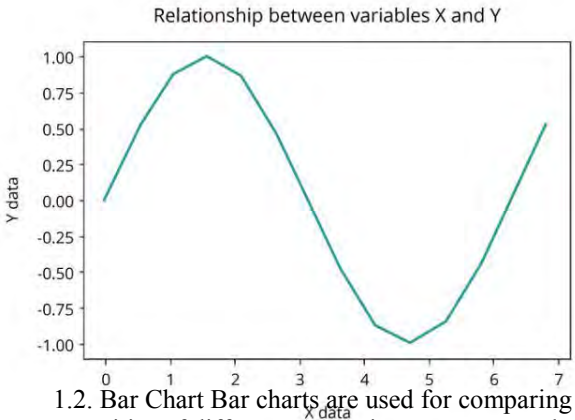
I. VISUALIZATON TECHNIQUE

Data visualization is applied in practically every field of knowledge. Scientists in various disciplines use computer techniques to model complex events and visualize phenomena that cannot be observed directly, such as weather patterns, medical conditions or mathematical relationships.

Data visualization provides an important suite of tools and techniques for gaining a qualitative understanding. The basic techniques are the following plots:

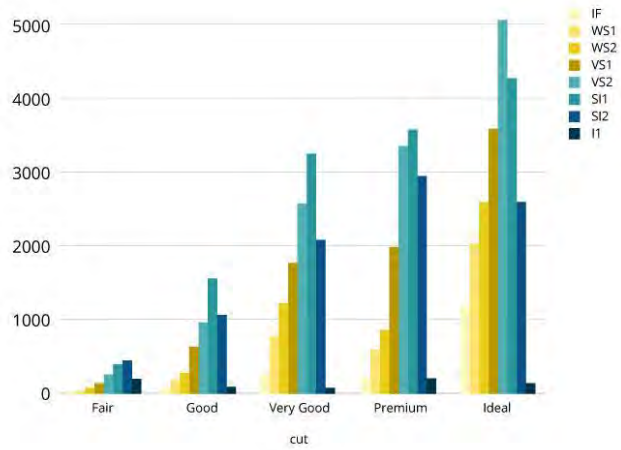
1.1. Line Plot

The simplest technique, a line plot is used to plot the relationship or dependence of one variable on another. To plot the relationship between the two variables, we can simply call the plot function.



1.2. Bar Chart Bar charts are used for comparing the quantities of different categories or groups. Values of a category are represented

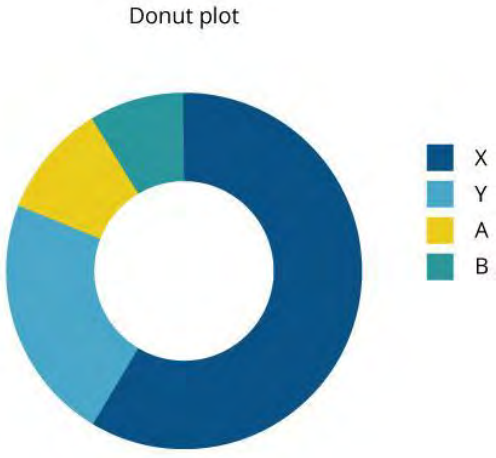
With the help of bars and they can be configured with vertical or horizontal bars, with the length or height of each



Bar representing the value.

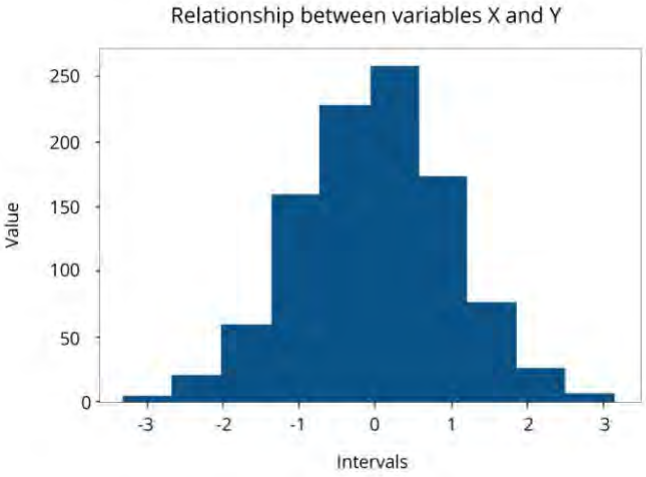
1.3 Pie and Donut Charts

There is much debate around the value of pie and donut charts. As a rule, they are used to compare the parts of a whole and are most effective when there are limited components and when text and percentages are included to describe the content. However, they can be difficult to interpret because the human eye has a hard time estimating areas and comparing visual angles.



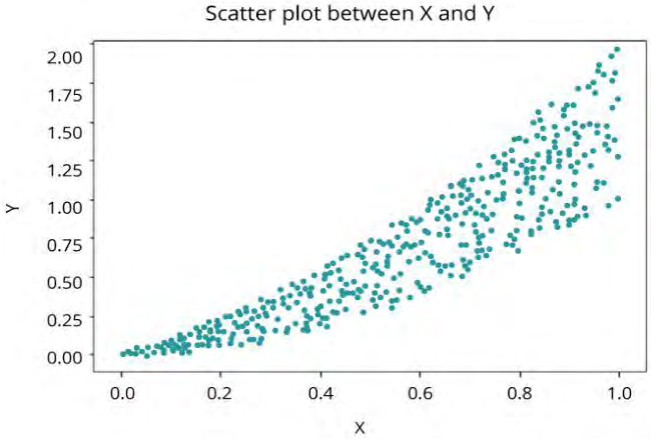
1.4 Histogram Plot

A histogram, representing the distribution of a continuous variable over a given interval or period of time, is one of the most frequently used data visualization techniques in machine learning. It plots the data by chunking it into intervals called 'bins'. It is used to inspect the underlying frequency distribution, outliers, skewness, and so on.



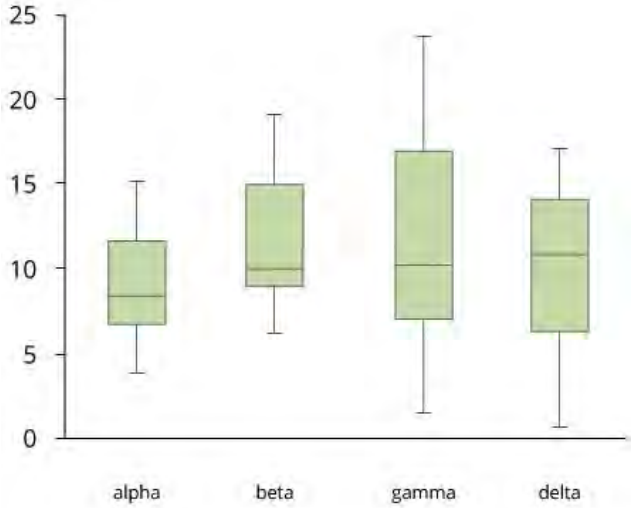
1.5 Scatter Plot

Another common visualization techniques is a scatter plot that is a two-dimensional plot representing the joint variation of two data items. Each marker (symbols such as dots, squares and plus signs) represents an observation. The marker position indicates the value for each observation. When you assign more than two measures, a scatter plot matrix is produced that is a series of scatter plots displaying every possible pairing of the measures that are assigned to the visualization. Scatter plots are used for examining the relationship, or correlations, between X and Y variables.



1.6 Box and Whisker Plot for Large Data

A binned box plot with whiskers shows the distribution of large data and easily see outliers. In its essence, it is a graphical display of five statistics (the minimum, lower quartile, median, upper quartile and maximum) that summarizes the distribution of a set of data. The lower quartile (25th percentile) is represented by the lower edge of the box, and the upper quartile (75th percentile) is represented by the upper edge of the box. The median (50th percentile) is represented by a central line that divides the box into sections. Extreme values are represented by whiskers that extend out from the edges of the box. Box plots are often used to understand the outliers in the data.



1.7 Word Clouds and Network Diagrams for Unstructured Data

The variety of big data brings challenges because semi structured and unstructured data require new visualization techniques. A word cloud visual represents the frequency of a word within a body of text with its relative size in the cloud. This technique is used on unstructured data as a way to display high- or low-frequency words.

Another visualization technique that can be used for semi structured or unstructured data is the network diagram. Network diagrams represent relationships as nodes (individual actors within the network) and ties (relationships between the individuals). They are used in many applications, for example for analysis of social networks or mapping product sales across geographic areas.

II. APPLICATION IN DATA ANALYTICS

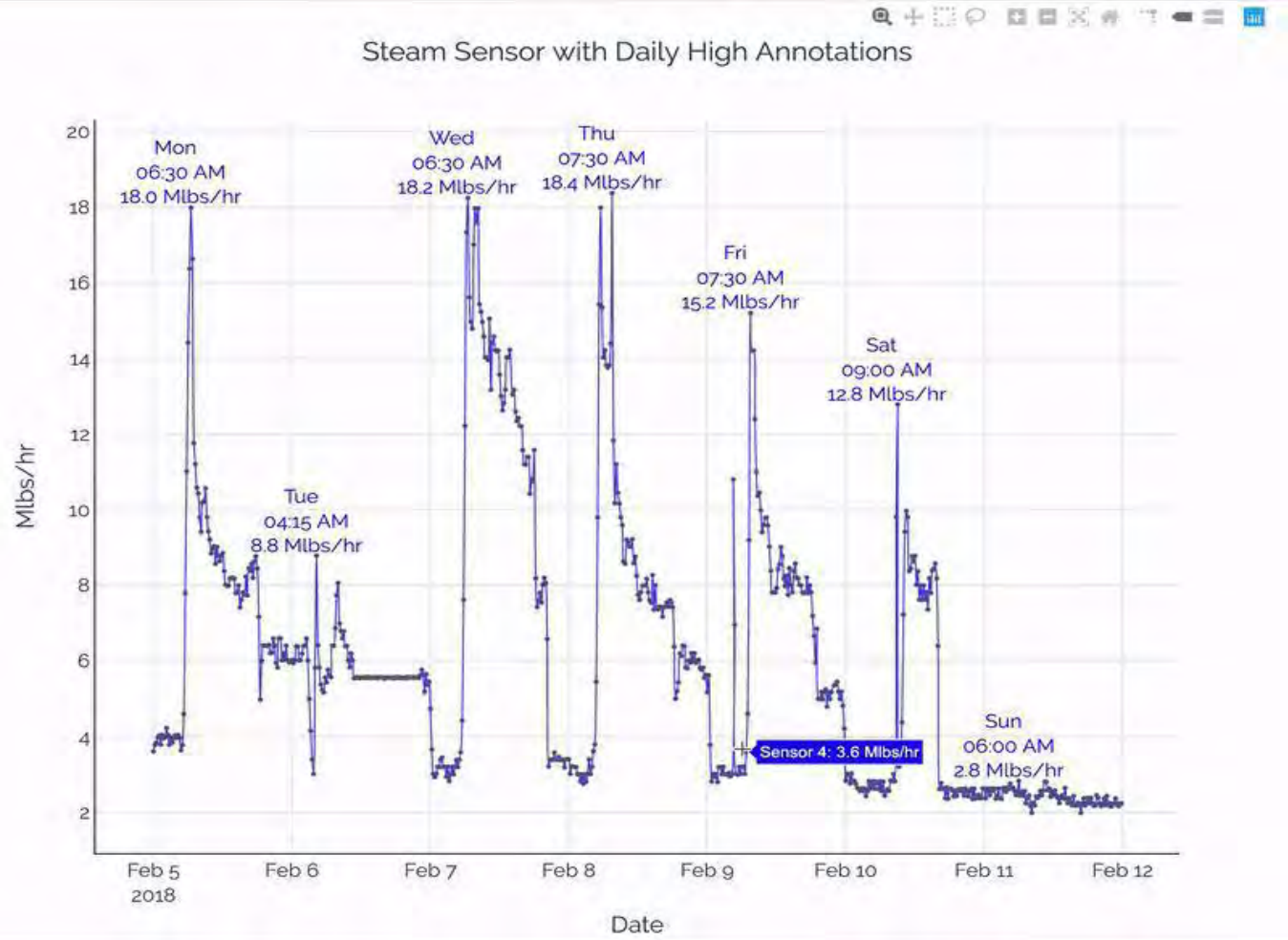
Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that

Might go undetected in text-based data closed and recognized easier with data visualization software.

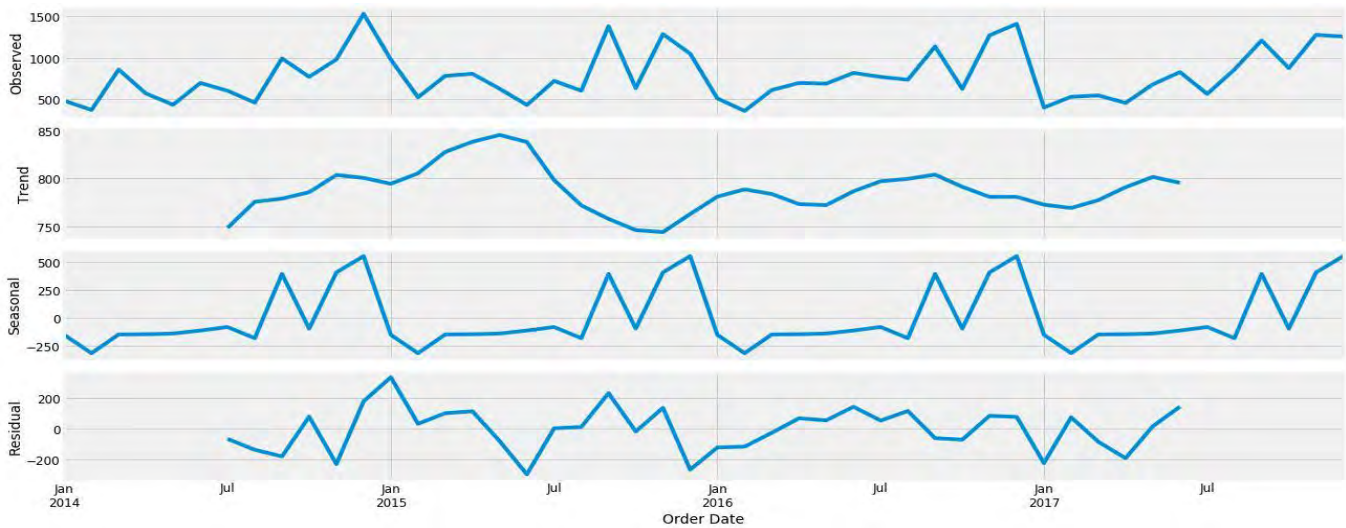
2.1 Forecasting

Data visualization play vital role in time series data where Graphs enable you to visualize many features of the data, including patterns, unusual observations, changes over time, and relationships between variables. Just as the type of data determines which forecasting method to use, it also determines which graphs are appropriate.

This type of interactive plots give deep insight about observations, trends, and seasonality and residual values



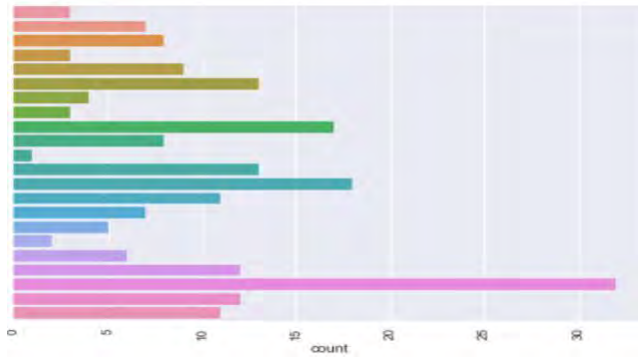
2.2 Public Health:



The ability to analyze and present data in an understandable manner is critical to the
Success of public health surveillance. Health researchers need useful and intelligent tools to aid their
Work [4]. Security is important in cloud-based medical data visualizations. Open any medical or health
Magazine today, and you will see all kinds of graphical representations.

2.3 Healthcare

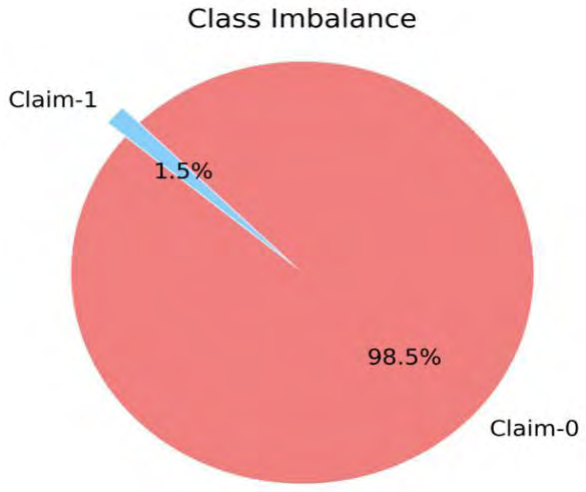
Visualization tools sometimes help in avoiding diagnosing error by eliminating human error in a higher risky environment and set a protective layer for patient safety. It creates a dashboard, i.e., e-health records and e-medical records for tracking and monitoring patient health records. Its computer-based recording system facilitates a better understanding of on-going medical operations.



Number of people effected in various area of country

2.4 Fraud Detection

We might come across the problem of class imbalance while solving classification problems like Anomaly Detection, fraud detection, disease diagnosis, etc. which can be easily viewed by various charts.



CHALLENGES

- Visual noise: Most of the objects in dataset are too relative to each other. Users cannot divide them as separate objects on the screen.
- Information loss: Reduction of visible data sets can be used, but leads to information loss.
- Large image perception: Data visualization methods are not only limited by aspect ratio and resolution of device, but also by physical perception limits.
- High rate of image change: Users observe data and cannot react to the number of data change or its intensity on display.
- High performance requirements: It can be hardly noticed in static visualization because of lower visualization speed requirements--high performance requirement.

IV. CONCLUSION

Visualizations can be static or dynamic. Interactive visualizations often lead to discovery and do a better job than static data tools. Interactive visualizations can help gain great insight from big data. Interactive brushing and linking between visualization approaches and networks or Web-based tools can facilitate the scientific process. Web-based visualization helps get dynamic data timely and keep visualizations up to date.

The extension of some conventional visualization approaches for data analysis is far from enough in functions. More new methods and tools of Data visualization should be developed for different Data analysis. Immersive virtual reality (VR) is a new and powerful method in handling high dimensionality and abstraction. It will facilitate Data visualization greatly.

V. Refernces

Genetic Algorithm for feature selection and parameter optimization to enhance learning on healthcare data

Abstract. **Classification algorithms as Support Vector Machine (SVM), and Neu-ral Network (NN) have provided considerably good results in diagnosis of critical care diseases. These Machine Learning algorithms have hyperparameters whose values if chosen optimally can provide enhanced learning. Operating on entire set of features is computationally expensive and requires more number of instances. Hence utilizing important features will reduce computation time. Both these objec-tives can simultaneously be obtained through nature inspired algorithms as Genetic Algorithm (GA). In the proposed work, GA has been utilized for feature selection as well as tuning the hyperparameters of SVM and NN. Optimal value of SVM Radial Basis Function (RBF) kernel parameters C and ɣ has been obtained. Simi-larly the novelty lies in identifying optimal no. of hidden layers, no. of hidden nodes, learning rate, momentum and optimizer for Multi Layer Perceptron (MLP) NN classifier. Results have been found better compared to utilizing Grid Search for the same. Further when reduced set of features are used for learning; sensitivity and precision score have found to be promising. Sensitivity is of more importance when healthcare is talked about and F1 score gives better accuracy since it is unbiased to data imbalance. Thus we have dealt with Multi objective optimization problem uti-lizing meta heuristics with improvement in diagnostic performance.**

Keywords: genetic algorithm, hyperparameter optimization, feature selection, sensitiv-ity.

I. INTRODUCTION

1.1 Support Vector Machine

In order to save sufficient computation time and build models that generalize well, penalty parameter C and kernel function parameter ɣ needs to be optimized. Though Grid algorithm can identify these optimal values through exhaustive search but cannot perform feature selection task. Support Vector Machine identifies the hyperplane D(x)=wx+b with maximum margin for linear separability as shown in Fig 1. For non-linear separable problem, kernel trick is applied as shown in Fig 2.

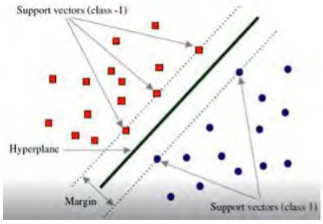


Fig. 1. Support Vector Machine

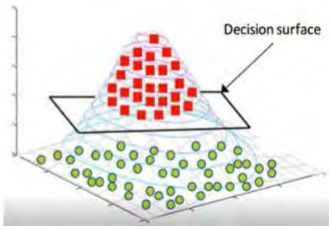


Fig. 2. Kernel Trick

1.2 Neural Network

Neural Network is modeled after the human brain and is capable of capturing complex non linear models. Activation functions operate in hidden nodes to transform net input to output. To avoid overfitting, bias is introduced. The activation function may be Lo-gistic / Sigmoid, ReLU (Rectified Linear Unit) etc. For a classification problem, output is the probability of belonging to a certain class. The model is retrained w.r.t. the error backpropagated until convergence. Thus we get optimal weights following Stochastic Gradient Descent (S.G.D.) algorithm shown in Fig. 3. An extension to S.G.D. is adam (adaptive moment estimation) optimization algorithm realizing the benefits of AdaGrad and RMSProp.

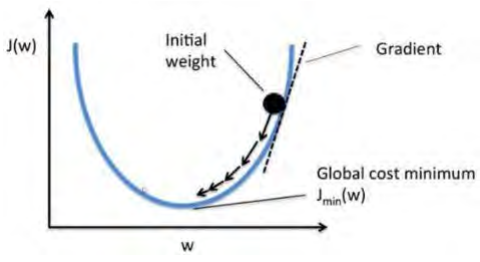


Fig. 3. Stochastic Gradient Descent

1.3 Genetic Algorithm

Genetic Algorithm are Evolutionary Algorithm inspired by natural selection, reproduc-tion and survival of fittest utilizing genetic operators – selection, crossover and muta-tion. Encoding is termed as a chromosome, a single bit being called Allele. It utilizes population of solutions in each iteration. It randomly selects solution to produce other solution and evaluates each potential solution. Two parents are chosen, genetic opera-tors are applied and moved to next generation until convergence.

Crossover is performed over random cutoff points. Mutation is bit flip with certain probability usually low. Elitism is to take best fitness value to next generation by re-placing the worst one. Fitness value is the value of objective function. Selection is through Roulette wheel where size is proportional to individual's fitness or Tournament Selection where several tournaments among few chromosomes are selected at random. Winner of these tournaments is selected for crossover.

II. LITERATURE SURVEY

Gupta S. et. al. have proposed that simultaneous optimization of hyperparameters of NN or SVM RBF kernel and feature subset selection through GA can yield high diag-nostic accuracy. Further they claim in another paper that feature selection is essential for simpler, faster, robust and more reliable ML models.

Huang et. al. has simultaneously optimized parameters and feature subset without compromising accuracy of SVM utilizing GA and compared performance with Grid search. Steps taken for GA based Feature selection are:

- Scaling to avoid attributes in greater numeric range dominate those in smaller nu-meric ranges
- Evaluate the fitness of each chromosome
- Look for better solutions by selection, crossover and mutation

The data is partitioned into training and independent test sets via kfold cross validation (CV). Fei et. al. searches best penalty and kernel function parameter and finally a less complex SVM model with fewer support vectors is proposed.

Bhatia et. al. has tried experimenting onto Cleveland multiclass using 'one against one' and binary class problem utilizing integer coded GA in order to avoid curse of dimensionality. The fitness value chosen is the classification accuracy achieved by fea-ture subset with 25 generation, 50 population, 0.8 crossover probability and 0.2 muta-tion probability. Similarly Arrythmia dataset with 17 classes received sensitivity of 91.40%, accuracy 98.99% and specificity 99.46% with genetic ensemble of classifiers by Plawiak et. al.

ANN is one of the most promising computational intelligence technique though the design requires setting of the structure and tuning complex parameter. Thus GA has been used by Ahmad et. al. to select significant features and identify optimal no. of hidden nodes along with Levenberg Marquardt (LM) backpropagation algorithm, the chromosome is shown in Fig. 4. Fitness function is calculated based on Mean Square Error (MSE) as shown in Eq. 1 and 2 below. MSE is converted to rank and chromosome with smaller MSE has better chance of selection. The author further utilized GA to fine tune weight of ANN. The chromosome being shown in Fig. 5.

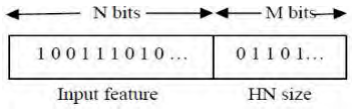


Fig. 4. Optimize features and hidden nodes [Ahmad et.al.]

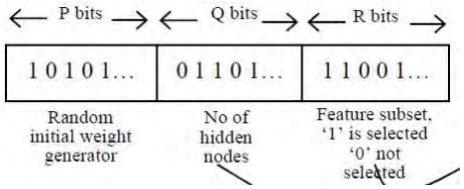


Fig. 5. Optimize initial weight, Hidden nodes, feature subset [Ahmad et.al.]

$$\text{MSE} \quad \frac{1}{0} \times \sum \sum (\quad - \quad)^2 \qquad (1)$$

$$= \quad \underline{0}$$

$$=1$$
$$=1$$

(2) $$fitness(i) = \frac{100 \times rank(i)}{\sum_{i=1}^{population\ size} rank(j)}$$

For RBF N/W, GA solves subset selection problem. Lacerda et. al. utilized holdout k fold CV and bootstrap function to accomplish the same. Zhao et. al. has also analyzed that inappropriate parameter settings lead to lower classification accuracy. Better per-formance may be achieved by discarding noisy, irrelevant and redundant features. The objective achieved was highest classification accuracy, least no. of selected features and least average processing time. Amin et. al. has utilized the global optimization ad-vantage of genetic algorithm for initialization of NN weights with faster, stable and accurate learning than Backpropagation (BP). The initialization of NN weights in BP algorithm is a blind process and there is slow convergence. LM optimization is fastest BP algorithm though requires more memory. Testing was done on 20 chromosomes in 100 generations with 12, 10 and 2 neurons in input, hidden and output layer respec-tively. The fitness function is based on MSE.

Jabbar et. al. identified the value of hyperparameter k in kNN with crossover prob=0.6 and mutation prob=0.033 for 20 chromosomes in 20 generation each resulted in enhanced accuracy for several diseased datasets. Pima Indian Diabetes dataset lies under normal distribution hence feature selection utilizing GA SVM enhanced accuracy and identified critical attributes by Santhanam et al. GA is used to select optimal set of features from 22 features of Parkinson's disease to 7 by adaboost and 10 by bagging ensemble utilizing fitness function as classification rate with 96.55% and 98.28% ac-curacy respectively by Fayyazifar et. al.

GA wrapper with embedded regularization achieved train and test accuracy of 98.83% and 93.61% respectively on Lung Cancer dataset by Liu et al. Increase in per-formance of NN by 10% through enhancing initial weights utilizing GA was obtained by Abrasadi et. al. Optimal feature subset with min. MSE and max. R^2 chosen with 22, 5 and 1 neuron in input, hidden and output layer respectively. Future work states opti-mizing learning rate and momentum factor. Gokulnath et. al. utilized Roulette wheel selection and SVM for classification with single point crossover followed by mutation to preserve genetic diversity and elitism. The data is preprocessed based on:

$$\text{Z-score normalisation} = \frac{(\ -\ )}{\quad} \quad \text{where } \mu\text{=mean and } \sigma = \text{s.d.} \qquad (3)$$

The fitness function f is based on classification accuracy f1 and no. of selected genes f2 as shown in Eq. 4-6 below. Comparison with other feature selection algorithms improved the classification accuracy by 4.64% onto Cleveland Heart Disease dataset.

$$\text{Classification accuracy of SVM } {}_1(\ ) = \frac{\quad +}{\quad + \ + \ +} \qquad (4)$$

$$\text{Number of selected genes } (\ ) = (1 - \frac{\quad}{2}\ ) \qquad (5)$$

$$\text{Fitness Function } (\ ) = {}_1(\ ) + (1 - \ ) {}_2(\ ) \qquad (6)$$

Similarly GA is able to optimize no. of base classifiers in an ensemble with signifi-cantly lower classification error and produce a model that is robust to outliers by both Fletcher et. al and Oh et. al. GA is also used to impute missing values when there is high rate of missing data with info gain as fitness function by Shahzad et. al.

III. EXPERIMENTATION AND RESULTS

Current work involves utilizing GA for parameter optimization of SVM and NN.
Genetic Algorithm is based on randomness. The methodology involved is as follows.
The same is reflected in the flowchart in Fig 6.

1. Set No. of iterations/generations (M), No. of individuals/chromosomes in initial population (N), crossover probability (p_c), mutation probability (p_m)
2. Create N population randomly of string size l
3. m=1, n=1
4. Select 2 parents through selection
5. Crossover the parents to get children at p_c
6. Mutate children at p_m
7. Calculate fitness of mutated children, save fitness value
8. Repeat 4-7 n/2 times (increment n)
9. Get a new generation of mutated children
10. Increment m
11. Repeat 4-10 M times
12. Chose best fitness value from last generation

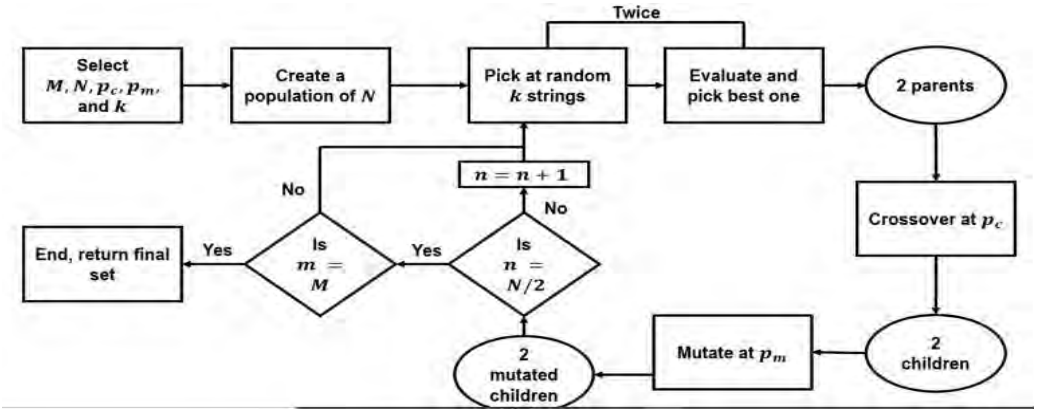


Fig. 6. Flowchart for applying GA

3.1 Dataset

According to yourtotalhealth website; WHO estimates 11.1 million deaths from Cardi-ovascular Artery Disease (CAD) in 2020. Further there are several datasets available on Kaggle.com on heart diseases. Cleveland and Stalog have less no. of instances whereas Hungarian and Switzerland have large missing values. Framingham Heart Dis-ease dataset predicting Ten Year Coronary Heart Disease (CHD) with 14428 instances and 15 features is preferred for experimentation due to availability of more data and less percentage of missing values. Any disease dataset has been found to be imbalanced since number of instances in healthy class will be less compared to diseased class.

Feature Engineering showed categorical features as gender, education level, cur-rently smoking, taking BP medication and whether stroke, hypertension, diabetes are prevalent. These features were transformed by creating dummy features and then de-leting redundant ones. Framingham heart disease dataset was containing unequal binary class distribution with 3596 and 644 instances respectively for patients suffering from TenYearCHD and healthy patients. Thus dataset was balanced through Synthetic mi-nority oversampling technique (SMOTE) which enhanced all parameters specially

Sen-sitivity results which are of more importance in medical diagnosis. Exploratory Data Analysis is done to reveal missing value in features BPMeds, education, glucose and totchol. Since the missing percentage was less, kNN imputation with Manhattan dis-tance for continuous and hamming for categorical was used for k=30. EDA also demon-strates features in differing scale. Standard Scaler is applied so that features in greater ranges do not overpower those in smaller ranges. Train:Test ratio kept is 60:40 as com-pared to 75:25. The table below reveals good sensitivity and F1 score for NN and SVM.

Accuracy=(TP+TN)/(P+N) (7)

Sensitivity or Recall = TP/P (8)

Specificity = TN/N (9)

Precision = TP/(TP+FP) (10)

F1 score = 2*Precision*Recall/(Precision+Recall) (11)

where TP=True Positive, TN=True negative, FP=False Positive, FN=False Negative P = TP+FN and N = FP + TN

Table 1. Comparison of base classifiers

Classifier	Accuracy (CV)	Sensitivity	Specificity	Precision	F1 score
Decision Tree	85 %	82 %	30 %	87 %	84 %
K Nearest Neighbor	85 %	71 %	49 %	88 %	79 %
MLP Neural Network	82 %	**84** %	28 %	87 %	**85** %
Support Vector Machine	83 %	**82** %	32 %	87 %	**84** %

3.2 GA for hyperparameter tuning SVM

The experimentation is performed for initial population of 100 chromosomes iterated over 50 generations for convergence. The validation set is chosen of 3 folds. The pa-rameters for GA are p_c=1, p_m=0.2, pop=50, gen=50, kfold=3. A stack of random chromosomes is created each time a new set of solution belongs to the population. Pre-cision is used to compute feature vector for binary chromosome. Then chromosome is decoded with respect to the range in which C and ɣ is allowable as shown in Eq. 7-8 below:

$$\text{Precision} = \frac{-}{2^{} - 1} \text{ where range (a,b), chromosome length=l} \quad (12)$$

$$\text{Decoding} = (\quad * 2) * + \quad (13)$$

The initial chromosome chosen is binary comprising of 1st 15 bits for ɣ and last 15 for C regularization hyperparameter of SVM RBF kernel respectively. The sample is shown in Fig. 7. Upper and Lower bound default ranges are b=1000, a=10 for C and b=0.99, a=0.05 for ɣ. This lower and upper bound has been provided through knowledge from published literature.

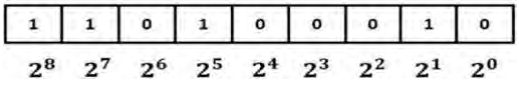


Fig. 7. Sample Chromosome of 9 bits

Decoding the chromosome/genotype to a value of C or ɣ is called phenotype. 3 contestants are chosen for tournament selection as per the knowledge gained from online sources. The basis is MSE with SVM classifier. Again this is the default number as stated in courses undergone. Diversity is created through unique random numbers.

Thus the minimization objective function = 1-Accuracy (14)

We find that the Accuracy for SVM w/o GA is **83**% and with is **84**%. Hyperparam-eter tuning with GA retrieved hyperparameter values as C=903.83 and ɣ=0.055. De-fault kernel parameters are C=1 and ɣ=0.1. Similar experimentation utilizing Grid search gave 80% accuracy with SVM linear kernel and C=1.

3.3 GA for hyperparameter tuning NN

Similar experimentation was performed with Multi Layer Perceptron (MLP) classi-fier to optimize weights used in training NN. The data has been shuffled for generali-zation. The no. of neurons range 6-10 and no. of hidden layers 3-8 based on similar experimentations done by other researchers. The chromosome is now 32 bit with 1^{st} 2 being combinatorial/discrete for no. of hidden layer and no. of neuron and next 30 being continuous for momentum γ and learning rate η respectively. A population of 50 chro-mosomes is generated for 50 generations. NN improves training speed and accuracy. Training finds appropriate value for weights and bias, utilizing BPNN keeping track of mutated child per generation and updating the weights computing increment by Eq 9-10 below:

Weight increment Δ = (∗)

$$\Delta = (* \_\_\_) + (* \Delta^{-1})$$

where η is the learning rate, γ is the momentum factor, Δ^{-1} is weight increment in previous iteration, and is weight gradient

The solver used is adam which gives better results than S.G.D. in terms of stability and speed. Upper and Lower bound default ranges are a=0.01, b=0.3 for momentum and a=0.01 and b=0.99 for learning rate based on literature reviewed. Two point cross-over has been chosen. The optimal value of hyperparameters obtained by GA is # of neurons=6, # of hidden layers=4, Momentum=0.086, Learning rate=0.060. Training accuracy achieved is **84** % with GA as compared to 82% w/o GA . Similar experimen-tation utilizing Grid search gave 80% accuracy with MLP NN utilizing activation=tanh, alpha=0.05, hidden_layer_sizes=(50, 50, 50), learning_rate being constant and solver as sgd.

3.4 GA for Feature Selection

The feature selection task when performed through NN with binary chromosome of size = # of features i.e. 15 reduces no. of features to 9 selecting male, age, education, currentSmoker, BPMeds, sysBP, diaBP, heartrate, glucose as high risk factors whereas feature selection task when performed through SVM returns 12 features i.e. male, age, education, cigsperday, BPMeds, prevStroke, diab, sysBP, diaBP, BMI, heartrate, glu-cose as high risk factors. Initial level of Feature Engineering through profiling showed correlation in features SysBP and DiaBP with Hypertension. Also correlation in Cur-rentSmoke and CigPerDay features was found through Spearman and Pearson Correla-tion. Results of feature selection through GA verify statistical findings. The experimen-tation is done for population of 80 chromosomes for 30 generations. 5 fold CV is per-formed for generalization based on courses undergone. Crossover is always performed whereas mutation is performed 20% of the time.

Table 2. Comparing performance with all and reduced features on SVM classifier

Classifier	Sensitivity	F1 score
With all features	82 %	84 %
With reduced features	89 %	88 %

Table 3. Comparing performance with all and reduced features on MLP NN classifier

Classifier	Sensitivity	F1 score
With all features	84 %	85 %
With reduced features	100 %	92 %

IV. CONCLUSION AND FUTURE SCOPE

The results claim the power of GA for parameter optimization of ML models and opti-mal feature selection utilizing MLP NN and SVM classifiers. Performance is enhanced further using Feature Engineering and Exploratory data analysis in the form of remov-ing data imbalance, imputing missing values, removing correlated features and treating categorical data. Research findings claim that though Grid search is used for hyperpa-rameter tuning but

cannot be used for feature selection and thus cannot fulfil multi-objective optimization..

Future work may test power of GA for finding best set of classifiers in an ensemble, find hyperparameter values for ensemble learners and impute large missing data as well. Current work can be extended utilizing NearMiss undersampling technique for balancing dataset and Multiple imputation with chained equations (MICE) for imputing missing values where multiple columns have missing data as in the case of Hungarian and Switzerland heart disease datasets. Currently the work is dealt with same number of neuron per layer for MLPNN, hence efforts shall be put in this direction in future work.

References

Arabasadi Z., Alizadehsani R., Roshanzamir M., Moosaei H., Yarifard A. A., "Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm, Elsevier ScienceDirect" Computer Methods and Programs in Biomedicine 141 (2017) 19–26.

Ahmad F., Mat-Isa N. A., Hussain Z., Boudville R., Osman M. K., "Genetic Algorithm - Artifi-cial Neural Network (GA-ANN) Hybrid Intelligence for Cancer Diagnosis, 2010 Second International Conference on Computational Intelligence, Communication Systems and Networks IEEE Computer Society.

Ahmad F., Mat Isa N. A., Halim M., Noor M., Hussain Z., "Intelligent Breast Cancer Diagnosis Using Hybrid GA-ANN", 2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks IEEE Computer Society.

Amin S. U., Agarwal K., Beg R., "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors, IEEE Conference on Information and Communication Technologies (ICT 2013) 1227-1331.

Bhatia S., Prakash P., Pillai G. N., "SVM Based Decision Support System for Heart Disease Classification with Integer-Coded Genetic Algorithm to Select Critical Features", WCECS Oc-tober 22 - 24, 2008, San Francisco, USA.

Fayyazifar N., Samadiani N., "Parkinson's Disease Detection Using Ensemble Techniques and Genetic Algorithm", IEEE Artificial Intellience and Signal Processing IEEE

Fei Y., Min H., "Simultaneous Feature with Support Vector Selection and Parameters Optimiza-tion using GA-Based SVM Solve the Binary Classification", 2016 First IEEE International Con-ference on Computer Communication and the Internet. Fletcher S., Verma B., Jan Z. M., Zhang M., "The Optimized Selection of Base-Classifiers for Ensemble Classification using a ith Applications 38 (2011) 5197–5204.

Multi-Objective Genetic Algorithm", 2018 IEEE International Joint Conference on Neural Networks (IJCNN).

Gokulnath C. B., Shantharajah S. P., "An optimized feature selection based on genetic approach and support vector machine for heart disease", Springer Science+Business Media, LLC, part of Springer Nature 16 Mar 2018.

Gupta S., Sedamkar R. R., "Apply Machine Learning for Healthcare to enhance performance and identify informative features", IEEE INDIACom; 6th International Conference on "Computing for Sustainable Global Development", 13th - 15th March, 2019, BVICAM, New Delhi (INDIA).

Gupta S., Sedamkar R. R., "Feature Selection to reduce dimensionality of heart disease dataset without compromising accuracy", 9th International Journal of Innovations in Engineering and Technology (IJIET) 2019.

Huang C. L., Wang C. J., "A GA-based feature selection and parameters optimization for support vector machines", Elsevier Expert Systems with Applications 31 (2006) 231–240.

Jabbar M. A., Deekshatulua B. L., Chandra P., "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm", Elsevier ScienceDirect Procedia Technology 10 (2013) 85
– 94.

Lacerda E. G. M. , Carvalho A. C. P. L. F. , Ludermir T. B. "A Study of Cross validation and Bootstrap as Objective Functions for Genetic Algorithms" IEEE Proceedings of the VII Brazilian Symposium on Neural Networks (SBRN'02).

Liu X. Y., Liang Y., Wang S., Yang Z. Y., Ye H. S., "A Hybrid Genetic Algorithm with Wrap-per-embedded Approaches for Feature Selection", IEEE Access 2018.

Oh D. Y., Gray J. B., "GA-Ensemble: a genetic algorithm for robust ensembles", Springer Mar 2013.

Plawiak P., "Novel genetic ensembles of classifiers applied to myocardium dysfunction recogni-tion based on ECG signals", Elsevier cienceDirect Swarm and Evolutionary Computation Base data 2017.

Santhanam T., Padmavathi M. S., "Application of K-Means and Genetic Algorithms for Dimen-sion Reduction by Integrating SVM for Diabetes Diagnosis", Elsevier ScienceDirect Procedia Computer Science 47 (2015) 76 – 83.

Shahzad W., Rehman Q., Ahmed E., "Missing Data Imputation using Genetic Algorithm for Supervised Learning", International Journal of Advanced Computer Science and Applications, Vol 8, No. 3, 2017.

Zhao M., Fu C., Ji L., Tang K., Zhou M., "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromo-somes", Elsevier Expert Systems w

Design and Development of a System for Predicting Threats Using Data Science

Gaurav Tripathi,Gaurav Rai,Sunny Singh
Computer Engineering Department
Thakur College of Engineering andTechnology
Mumbai, India
tripathigaurav150@gmail.com,raiigauravv@gmail.com,Singh.sunny0852@gmail.com

Abstract—**ImpactofSocialMediaonahumanbeingis moresevereanddangerous ratherthantheirenemyas we arelivingin aworldwhere socialmediais the second mostessentialpartofourlife,likewithfood andwater humanscannotlive,samecanbesaidabout thesocial mediaasitconnectspeoplewhoarelivingfaraway to thosewhoareclosetoyou.Somakingsocialmediasafe heavenisadifficultjobbutat-leasttakingameasure canbedonebeforeanychaos takesplaceduetosocial mediaactivitycouldbedone. Associal mediadoesn't boundbyany boundaries ofcountriesandgovernment rules even sometimesinternationalrulesarealsonot enoughtohandlethethreats,government havemany systemsbutnotasintegratedonetohandle thepublic levelthreatswhichhadalreadytakenplacein countries like India.Themainaimofthisresearchistoutilize natural languageprocessing,dataanalytics,bigdata, machinelearningideasfortheacknowledgement of upcomingthreatsandpredictionofsocialmediathrea ts dependent onthereal-time scenarios. Here inthis paper,unsupervisednaturallanguageprocessing is appliedtosocialmediadata,withareasonto predict, identifythepeopleandsource, responsible for spreadingthreats.**

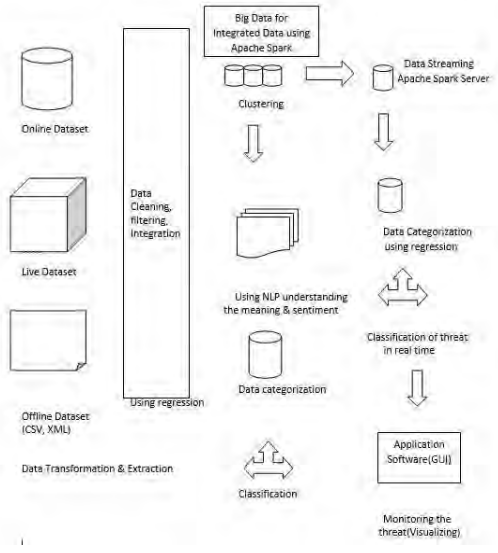
Keywords— Socialmediathreats analysis; Threatprediction of social media; Natural LanguageProcessing;Machine Learning; DataAnalytics;Big Data;Convolutional Neural Network ;

I. INTRODUCTION

Over thelastfewyears, onlinecommunicationhas movedtowardsocial-driventechnologies, likesocialmedia networkssuchas WhatsApp,Facebook,Instagram,Twitter andothersocialmediaplatformssuchas bloggingwebsites, YouTube Videos,onlinevirtualcommunities,andonline petitionplatformssuchas change.org, etc. Thesesocial technologies havestarted a revolution in user-generated data, online global communities, and rich human behavior-relatedcontent.Understanding humanpreferences are importantto the development of applications andplatformsforpredictingthreatsfromsocialmediadata using datascience.Thispaperdiscussesthe role of social media datatounderstandthebehaviorofhumansregarding their onlineactivitybasedondata.

Withtheadvancementinthelatesttechnology about sentimentanalysisandpredictiveanalytics,ithas opened manyavenuesforresearchersandenterprisestoundersta nd thehumanmentalstateandtheirsocialbehaviorina better way.Theproposedchallengeis topredictpeople'sactivities onsocialmedia,tohelpineliminatingandreducing the percentageofanyincident(Stampede,Mob-lynching,etc.) before it'soccurring. There areseveralexistingsystems whicharecurrentlyusedbythegovernmentof Indiathey are listed as NETRA(NetworkTrafficAnalysisSystem), CCTNS(Crime and Criminal Tracking &Networks System), Social MediaLabsandwhereNETRAismost effective nowadays as it is used byintelligencebureau, India's domestic intelligence agency and research & analysis wing(RAW) asNETRAcananalyze voice traffic passingthrough severalsocialmediaplatformsandcan interceptmessageswithsuchkeywordlike'attack','bomb','kill'inreal-timesystemfromtheenormous amountof tweets, statusupdates,blogs,forums.

Thepurposeistoaimatmaking useofunsupervised natural language processing and machine learning algorithms(Clustering,LinearandLogisticregression)a nd datascience (Data Pre-processing, Big Data, hybrid algorithm) ininterpreting SocialWebDataofvarious platformslikeFacebook,Twitter,Googletrendsand other



themintosixcategoriessuchasToxic,Severe Toxic, Obscene,Insult,Threat,IdentityHatetopredictthe threat andtoxicitylevelofthatparticulartrendingtopic.

II. RELATEDWORK

A. Overview

Sonowadaystherearemany tools and technology availableinthemarketwhich canbeused toanalyze the socialmediadatabutmostofitareusing formarketing purposethattooeven arenot cost-effective**,** withthe advancesintechnologyaboutsentimentanalysis and predictiveanalytics,ithasopened manyavenuesfor researchersandenterprisestounderstandthehumanment al statebetter. Theproposedchallengeistopredictpeople's activities onsocial media, to help in eliminating and reducingthepercentageofanyincidentbeforeitoccurs. The purposeis aimedat makinguseunsupervisednatural languageprocessingandmachinelearning algorithms (Clustering, Linearand Logistic regression) and data science (Data Pre-processing, Big Data, hybrid algorithm)ininterpreting SocialWebDataofvarious platformslikeFacebook,Twitter,Googletrendsand other varioussourcesofmachinelearningalgorithms(Clusteri ng, Linear and Logistic regression) and data science (Data Pre-processing,BigData)ininterpretingSocial WebData of various platforms like Facebook, Twitter, Google trendsandothervarious sources. Social Media Analysis will bemonitoring, collecting,andanalyzingthedataofa Facebook page (posts, comments, likes, shares) and a Twitter profile (tweets, retweets, mentions, and public tweetscontainingone/twokeywords only) andfromother various sources for predicting the future or upcoming threats.

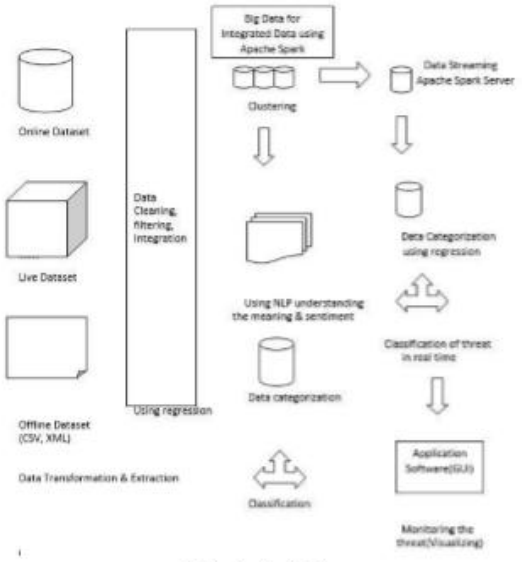


Fig I. Architecture

B. Algorithm

NLTKClassifier-NLTKClassifiertopredict thesentiment ofthecommentsandtweetsthatareeithertheyare positive ornegativebyusingacorpusdatasettoclusterthedatainto categories.
Deep Learning Unsupervised Threat Classification

C. Application

FigII. Flowchart

algorithmusingCNN andWord2vec.
CNNisaclassofdeep,feed-forwardartificial neural networks(whereconnectionsbetweennodesdonotforma cycle)&useavariationofmultilayerperceptronsdesigned torequireminimalpreprocessing.
Word2vec-isagroupofrelatedmodelsthatareusedto produce wordembeddings. These models are shallow, two-layer neural networks thataretrainedtoreconstruct linguisticcontextsofwords.
Pytrendsislibrarywhichisasubstitutetogoogle trends libraryforpythontoextract the data from google,youtube, imagesfortrendingtopicsandpastinformationrelateddat a whichwere searchedongoogle.

- TopredicttheReal-timethreatwhichistrending onsocialmediainterms ofthehashtag.
- Individually trackadetailedreviewofaperson whatkindoftweets andcommentstheyusuallydo.

- StreamingReal-timedatafromvarioussourcesand analyzingthemforpredictingthethreat.
- Real-timeclassificationofthreatintermsoflow severehighdependonthethresholdandpopulation oftheregion
- Region-wiseclassificationofhashtagsandtweets and furtheranalyzing the particular hashtag to predictwhatkindoftweets are trending.

D.ExpectedOutcome

This project upon completion of full functionalities by having access to all social media platforms will easily predictthethreat,fakenewsandcategorizing thetrending hashtagintermsofobscene, toxic, severetoxic, threat, identity hate, insultby analyzingeverycommentsand tweetbasedontheregion,thismighthelptocategorizethe trending hashtag or any trendingtopicwhethertheyare worthyofsharing,re-tweets,comments,etc.AsimpleGUI will help to displaythe real-time threat analysis by streaming data fromvarious social mediaplatformto checkthethreatlevelbasedontheregioninthreecategorie s low,moderateandseverebasedonthepopulation ofthat region.

II. METHODOLOGY

Afterthemethodologybeingretrievinguser'sdatafrom various socialmedia platforms instructured/unstructured formatmostprobablyinJSONformatusing theAPIs provided by these platforms. The extracteddata isthen cleanedandfilteredforourspecificrequirements. Thedata thenbasicallystreamedtotheclient-sideapplication using theSPARKStreamingandSocketStream programwiththe helpofFlask.FlaskisalightweightWSGI webapplication framework.Itisdesignedtomakegettingstartedquickand easy,withtheabilitytoscaleuptocomplexapplications.It beganasasimplewrapperaround WerkzeugandJinjaand hasbecomeoneofthemostpopularPythonwebapplicatio n frameworks.Themainmethodology forThreatAnalysisis theDeeplearningNLPClassificationUnsupervised model usingconvolution2DCNNfortextclassificationmethodi s usedtotrainthedataset.Andusing NLTKCorpus datafor sentimentanalysisfortherangeofsentimentsofpeople in termsofpositiveandnegative statements.Everythreatis classified in three categories low, moderate and severe basedonclassificationofevery tweetand comments into several terms like Toxic, SevereToxic,Obscene,Insult, Threat,IdentityHateandbycounting theseandevery trendingcanbetaggedwhetheritisconsideredtore-tweet, repost,andcomment.

III. TESTINGENVIRONMENT

In our research, we focus on a collection ofdatafrom varioussocialmediaplatformssuchasFacebook, Twitter, and various other sources, In comparison with previous work,thispaperfocusesonmoreintegratedretrievalofda ta andreal-timeanalyticsandpredictiononthestreamingdata ina batch.

At first, thesubjectofthispaperis thedetectionofhashtags, comments,trendingtopicsonseveralsocialmediaplatfor ms and posts whichareposingthreatsintermsof "toxic", "insult","identity hate","obscene", "threat",“identity hate” tosocietyinrealtime .

Secondarily,inthispaper, weutilized the30% datasetof toxic comment challenge and whichwerepublishedon Kaggleand70%ofdatafromvariousotherplatformsbyca tegorizing them individually into threat category of several languages as every state in indiahas different contextofwritingthetext, toidentifytheabove-mentioned terms ofthecomments,posts,andtweets thismayhelpusto achieveandidentifythehashtags, trending topicsandpost whichareposingthreatonindividual,society,andbusines s inreal-time.

IV. CONCLUSION

ThreatAnalysisofSocialMediaDataofvarious platforms willbehelpfultosocietyas socialmediais thesecondmost important aspect of humanbeingsomakingitsafeisa difficulttaskbutsomehowa particularsecurityandsafetyin termsofsecuritymeasurescanbeprovided byachieving certainterms liketrendinghashtags,trendingtopicsinterms ofparticularcategoriessuchastoxic,severetoxic,obscen e, insult,threat,identityhatewhichcould prevent usersfrom commenting,tweeting,re-tweeting,repostingsomanyof thethreatscanbeavoidedjustbytaggingthemas thesewill beimplementedregion-wiseifit makeseasier, efficientand usingreal-timepredictionitcanbecome one ofthemost accurate,efficientandeffectivewaystoprovide securityto societyandhelpreinforcementdepartmentsofcountriesa nd helppeopletounderstandtheperspective andobjective of trollers,criminals,fakenewsagencies and political interferences. This project providesasimpleGUIwhich makeseasiertopredictandanalyzethe threat of various platformindividuallyandpredictthethreataswell. This systemisanintegrated,real-time dataprocessingofvarious social mediaplatforms in anefficientwaytodetectthe threatswhichare posedbysocialmedia.

V. ACKNOWLEDGMENT

WewishtoexpresssincerethankstoMr. VijayJainand Dr.AnandKhandareforhelpinginour researchand our ComputerEngineeringdepartmentforarranging sucha

platform.sothatwecanresearchallabout SocialMedia Analyticsfromseveralplatforms. Wealsoextend our heartfelt thanks to our colleagues, family members,and well-wishers.

References

[1] MichaelFire,RoyGoldschmidt,andYuvalElovici2014. Online Social Networks: Threats andSolutions(2014)

[2] Jennifer Golbeck, Zahra Ashktorab, Rashad O.Banjo, Alexandra Berlinger, SiddharthBhagwan(2017,ALargeHuman-LabeledCorpusforOnlineHarassmentResearch(2017).

[3] MiltiadisKandias, VasilisStavrou,NickBozovic, DimitrisGritzalis2013.Proactive InsiderThreat Detection Through Social Media: The YouTube Case (2013)

[4] RachaAjami, Noha Ramadan, Nader Mohamed, Jameela Al-Jaroodi(Security Challenges and Approaches in Online Social Networks 2011)

[5] Yakshi Sharma Dept of I.T, U.I.E.T Panjab University, Chandigarh, 160014 India Veenu Mangat2015. Mandeep Kaur practical approach to Sentiment Analysis of Hindi tweets(2015)

[6] Aristidis G. Vrahatis, Sotiris K. Tasoulis,VassilisP. Plagianakos, Aristidis G. Vrahatis 2018.Convolutional Neural Networks for Toxic Comment Classification (2018)

[7] WeiminLuo, JingboLiu,JingLiu2009.AnAnalysis of Security in Social Networks (2009)

[8] Soumya.T.R, S. Revathy 2018. SurveyonThreatsin Online Social Media (2018)

[9] YojanaGoyal, Anand Sharma 2019. A Semantic Approach for Cyber Threat Prediction Using Machine Learning(2019)

[10] AnniSapountzi, Kostas E. Psannis (Social Networking Data Analysis Tools & Challenges 2016)

[11] HarriJalonen (Negative emotions in social mediaas a managerial challenge 2014)

[12] Natalya (Natalie) Bazarova, Yoon Hyung Choi, Victoria SchwandaSosik, Dan Cosley (Social Sharing of Emotions on Facebook 2015)

System for Authentication Using Voice Biometric

Saurabh Pandey, Tarun Pandey, Smriti Singh
Department of Computer Engineering
Thakur college of Engineering and Technology
Mumbai, India
Sp157311@gmail.com, tarunpandey1998@gmail.com, smrititanwar07@gmail.com

Abstract— **traditionally the use of biometric devices has improved our ability to provide authenticated access to physical installations. With the rise in security breaches in recent years, many organizations believed that two- factor authentication would be the answer. [4]Voice Biometric can be one of the factors of authentication. In this paper, there will be three phases- Feature Extraction, Training and Matching. The features will be extracted by using MFCC (Mel Frequency Campestral Co-efficient) technique. In the training phase the extracted features will be trained using the Gaussian Mixture Model (GMM). In the last phase, the algorithm on a dataset and the model will be used to predict the appropriate classes by finding the log likelihood of voice sample.**

Keywords— MFCC- Mel-frequency Cepstral Coefficients GMM-Gaussian Mixture ModelEM- Expectation Maximization

I. INTRODUCTION

With recent advancement in technology voice recognition has become one of the efficient measure that is used to provide protection to human's computerized and electronic belongings. . It is one of the types of biometric that is used to identify and authenticate user on the basis of his/her voice. [2]Voice recognition is divided into two types: text dependent and text independent. Text dependent recognition identifies user against a phrase while text independent recognition identifies the user irrespective of what he is saying. The success in both cases depends upon the various speaker characteristics which differentiate the one speaker from other. All voice recognition systems comprises of two modules, feature extraction and feature matching. In feature extraction, data from the voice sample is extracted to represent the certain speaker and in feature matching the extracted features from the input voice sample is matched against a set known speaker. Speaker recognition methods can be divided into text-independent and text-dependent methods. In a text-independent system, speaker models capture characteristics of somebody's speech which show up irrespective of what one is saying. In a text-dependent system, on the other hand, the recognition of the speaker's identity is based on his or her speaking on one or more specific phrases, like passwords, card numbers, PIN codes, etc. Every technology of speaker recognition, identification and verification, whether text- independent and text dependent, each has its own advantages and disadvantages and may require different treatments and techniques. The choice of which technology to use is application-specific. At the highest level, all. Speaker recognition systems contain two main modules feature extraction and feature matching

II. EASE OF USE

A. Entertainment:

Voice recognition can be used to change TV or radio channels, open and close screens, and play movies. It can also help personalize customer experience. For instance, services such as Netflix and Hula can be personalized by determining the age of the user through voice analysis, enabling them to access age-appropriate content.

B . Healthcare:

The global healthcare biometric market is expected to reach USD 14.5 billion by 2025, according to a recent report by Grand View Research, Inc. In an industry where data security is paramount, physicians can use voice biometrics to dictate and record patient's health conditions directly into the system and securely retrieve patient's personal history. This can significantly benefit patients who need to share medical records between various doctors. The system can also help dramatically reduce fraud for providers and payers by automating payment collection, and improve patient satisfaction by offering an additional payment option.

C. Banking:

Customers can use voice authentication to operate bank lockers. Banks, on the other hand, can leverage the system to enable highly secure and advanced voice-based payments. With fraud on the rise, credit card companies and banks such as Citibank and ANZ use voice biometrics to proactively identify fraudsters and authenticate callers at their call

centers.*D. Education:* Educational institutions can use voice recognition to provide flexibility to students with visual disability, helping them take online exams using voice authentication.

E. Independent Software Vendors:

For ISVs, voice authentication can enable enterprise sign on mechanisms such as those based on Active Directory, enabling authentication uniformity across enterprise applications and strengthening compliance with accessibility standards.

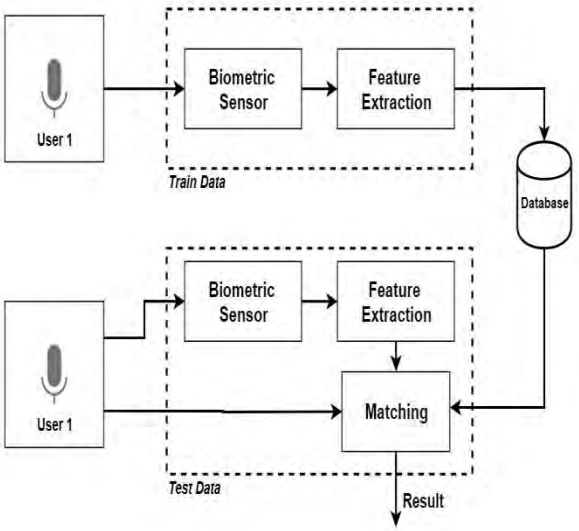


Fig. 1: Architecture

III. HOW IT WORKS

The first step in any voice recognition system is for the user to give an input by speaking a word or a phrase into a microphone as shown in Fig. 1. Then an analog to digital converter converts the electrical signal into digitized form and stored it in the memory. The computer then attempts to determine the meaning of a voice sample by matching it with a template that has a known meaning. This is an analogy to the traditional inputs from a keyboard. The greatest hurdles in Speaker- independent speech recognition systems are articulations and variety of accents used by the people having different nationalities. Other factors that present a challenge to voice recognition technology are acoustical noise and variations in recording environment which are beyond speaker variability. The developed system is consisting of three processes:

- Features extracting
- Training
- Matching

In first process, the developed system will result as computed features of human voice. These features are voice features which are taken from the persons. These features are extracted by using MFCC (Mel Frequency Campestral Coefficient) technique. [1] MFCC is used as the acoustic features of human voice. It considers the human voice pitch in the form of frequencies and scale them on the Mel scale, these extracted features are unique to others. In training process, the extracted features are trained using the Gaussian Mixture Modeling. Expectation Maximization (EM) algorithm is used to train the extracted features of human voice in system and then finally used to store in database. In the last phase, the model so constructed will be used in predicting the appropriate class using the validation dataset. We implement MFCC with GMM techniques in order to identify the speaker.

A. Abbreviations and Acronyms

MFCC: Mel Frequency Cepstral Coefficients
GMM: Gaussian Mixture Model
Expectation Maximization Units

B. Algorithms

1) MFCC Algorithm:

Mel Frequency Cepstral Coefficients (MFCCs) are a feature widely used in automatic speech and speaker recognition. They were introduced by Davis and Mermelstein in the 1980's, and have been state-of-the-art ever since. Prior to the introduction of MFCCs, Linear Prediction Coefficients (LPCs) and Linear Prediction Cepstral Coefficients (LPCCs) (click here for a tutorial on cepstrum and LPCCs) and were the main feature type for automatic speech recognition (ASR), especially with HMM classifiers. [6] The Mel scale relates perceived frequency, or pitch, of a pure tone to its actual measured frequency. Humans are much better at discerning small changes in pitch at low frequencies than they are at high frequencies. Incorporating this scale makes our features match closely to what humans hear.

2) GMM Algorithm:

In training process, [3] the extracted features are trained using the Gaussian Mixture Modeling. It recognizes the speaker on the basis of log probability of voice vector and compares it to previously stored value. The log probability, equal to the stored value provides access to the entire speaker. Suppose there are K clusters (For the sake of simplicity here it is assumed that the number of clusters is known and it is K). So mu and sum is also

estimated for each k. Had it been only one distribution, they would have been estimated by maximum-likelihood method. But since there are K such clusters and the probability density is defined as a linear function of densities of all these K distributions.

3) Expectation-Maximization

It is used to estimate the maximum likelihood and it returns the mean Variance Weight W and log Probability. The essence of Expectation-Maximization algorithm is to use the available observed data of the dataset to estimate the missing data and then using that data to update the values of the parameters.

C. Applications

Entertainment
Healthcare
Banking
Education
Independent software vendors (ISV)

D. Advantages

Increased Security
Decreased Fraud
Improved Customer Experience
Reduced Costs
High Reliability Rate
Easy to use
Minimally Invasive

E. Methodology

The proposed system was designed using methodology of incremental model. Incremental Model is used for designing, integrating and testing the system. Incremental Model is chosen for this system because system can be developed and delivered in increments, accommodate changes that evolve with time, are easy to test and debug and easier to manage risks involved.

F. Objectives

- To makes a system that identifies a person by processing his/her voice.
- To make system more secure by preventing unauthorized access.
- Performance Evaluation of proposed framework by comparing the performance of existing Voice Recognition System.
- Design of a Security Template to secure Voice recognition system
- To develop or improve existing algorithms to make the Voice recognition accurate on Voice with background noise. And also increase the security of a system to a certain extent when combined with another security step.

IVR (Interactive Voice Response) systems as shown if Fig.2, can be built to interact with customers and the voice print can be used to authenticate the customer and progress in the task which he/she wants to do. These systems can be used in multiple domains for example, telecommunication domain for verification of the user, banking domain for a customer to check balance, change pin or password, withdraw money, social media websites to login, etc.. This system can be used in any domain where security is a concern.

All of us are aware of the word 'jackpotting' which is nothing but illegally withdrawing money from an ATM. An ATM can be built with a feature of voice recognition so that only authorized customers can do the transactions safely and securely.

IV. FUTURE SCOPE

This paper is on how voice authentication can be used in day to day life to make systems more secured because security is a very important thing in today's world with everything going digital. A system is highly vulnerable to impersonation attacks i.e. authentication and authorization attacks. Impersonation means gaining access to resources without the correct credentials. System cannot figure out who the user is until and unless correct username match to correct password. Bringing biometrics in picture reduces a burden of task like remembering credentials and entering whenever it is needed. Voice authentication will reduce time

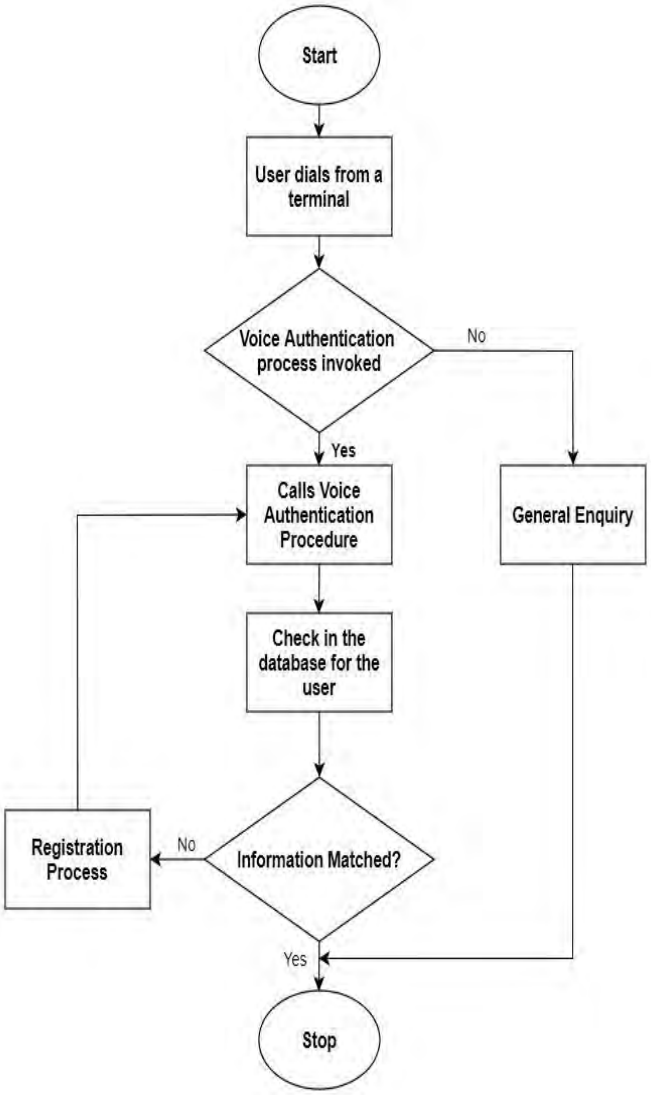


Fig. 2: IVR Flowchart. IMPACT ANALYSIS

Current solutions for passwords and authentication are insecure and have been hacked very regularly. Biometric based authentication provides a promising solution because voice biometrics, fingerprints, creates unique identifiers for individuals [9]. With the rise in security breaches in recent years, many organizations believed that two-factor authentication would be the answer. Voice Biometric can be one of the factors of authentication. A more commonly overlooked advantage of voice biometric solutions is that they actually hold the potential to significantly improve customer experiences. With voice biometrics products, callers no longer need to provide passcodes or PINs or provide answers to challenge questions in order to verify their identity. This makes voice biometrics ideal for various Omni channel and multichannel deployments, because once a customer is enrolled; his or her voiceprint can be leveraged across all of your company's support channels. This seamless experience not only makes the process easier and more efficient for your customer, but it also leads to improved CSAT, NPS, and Customer Effort Scores, as highlighted by recent Forrester research. In fact, it has been found that voice biometrics can reduce the time it takes to verify a caller's identity from. This advantage offers the ability to dramatically improve not only call personalization, but also customer experiences. There is many different ways that voice and speech recognition is helping commercial organizations. Many business concerns that want to excel at customer service use this 'cutting-edge' technology to enhance their customer interaction experience and reduce organizational expenses. Since voice authentication technology can take the place of living agents, it can also eliminate all problems, such as unauthorized access, theft of confidential documents, fraud etc. Through Voice Recognition technology, all calls can be handled by the system rather than a receptionist. This ensures a seamlessly uniform experience. Furthermore, the customers will be fully satisfied with the quality of the service. Voice authentication is a tremendous boon for organizations which will help them secure their confidential data from unauthorized access. Moreover, instead of remembering the pin code of bank account, customers can easily access their account and carry out their activities effortlessly and all of this will

increase customer satisfaction as they will get rid from the burden of remembering pin codes of their bank accounts.

V. CONCLUSIONS

In this paper, efforts have been made to overcome existing security threats by implementing two factor voice authentications. Limitation of uncontrolled environment conditions related to authentication using voice recognition has been taken care of in this paper. MFCC features were used which solved the problem of unfavourable human conditions such as sore throat as it models the vocal tract, tongue and teeth which affects the voice of an individual.

VI. ACKNOWLEDGMENT

We sincerely thank our guide Ms. Veena Kulkarni for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work. We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

References

[1] Tahira Mahboob, Memoona Khanum, Malik Sikkandar Hayat Khiyal and Ruqia Bibi, "*Speaker Identification Using GMM with MFCC*", IEEE Trans on Computational Intelligence and Computing Research, Vol. 12, No.2, pp. 978-1988, March 2015.

[2] Annie Shoup, Tanya Talkar, Jodie Chen, Anubhav Jain." *An Overview and Analysis of Voice Authentication Methods*", International Journal of Computer Science Issues, Vol. 10, Issue 2, pp. 1026-1035, 2016.

[3] G. Suvarna Kumar, Raju, K.A. Prasad, "*Speaker Recognition using GMM*", International Journal of Engineering Science and Technology, Vol. 2(6), pp. 2428-243,2010

[4] J. Bonneau et. al., "*The Quest To Replace Passwords*", IEEE Symposium on Security and Privacy, pp. 553-567,2012

[5] Krishna Dharavath, F.A. Talukdar and R.H. Laskar, "*The need of Authentication and issues related to privacy*", IEEE International Conference on Computational Intelligence and Computing Research, pp. 978-985, 2013

[6] "Mel-Frequency Cepstrum", https://en.wikipedia.org/wiki/Mel-frequency cepstrum (August 5, 2019).

[7] "python_speech_features", https://github.com/jameslyons/python speech features (August 10, 2019)

[8] Jain, A.K.Ross, A.Prabhakar, "*An introduction to biometric recognition*", Vol. 14, Issue 1, pp. 4-20, Jan. 2004.

[9] Jain, A.K. Ross, A. Pankanti, S., "*Biometrics: a tool for information security*" Vol. 1, Issue 2, pp.125-143, June 2006.

[10] D. Reynolds and R. Rose, "*Robust text-independent speaker identification using Gaussian Mixture speaker models*", IEEE Trans on Speech Audio, Vol. 3, No. 1, pp. 72-83, Jan.1995

Customer Relationship Management using Data Mining techniques

Shukla anurag,Jha Nilesh,Mohd Salim,Prof. Rashmi Thakur
Computer Engineering, TCET
University of Mumbai
shuklaanurag998@gmail.com,nilesh.jha412@gmail.com,shaikhsalim3105@gmail.com,rashmi.thakur@thakureducation.org

ABSTRACT-

Customer relationship management (CRM) is an approach to manage a company's interaction with current and potential customers. It utilizes information examination about clients' history with an organization to improve business associations with clients, explicitly concentrating on client maintenance and at last driving deals development.

Customer relationship management has many components such as Sales Force automation, Human Resources management, Customer Service, Business reporting, Analytics etc. In this paper, we will be discussing our approach for building various tools used for Customer relationship management.

INTRODUCTION

In a Business to Business atmosphere, suppliers and service providers typically ought to perceive the characteristics of their customers. Customer attraction and satisfaction are the main objectives of any leading company. This paper discusses the methods used for building the tools which help in CRM tasks like Customer Loyalty prediction, Market Basket Analysis, Feedback Chabot etc.

RFM analysis

RFM analysis1 has been utilized in direct marketing for many decades (Baier, Ruff, & Chakra borty, 2002).

This system identifies client behavior and represents client behavior characteristics by 3 variables as follows:

(1) Regency of the last purchase which refers to the interval between the time that the latest consuming behavior happens and present.

(2) Frequency of the purchases which refers to the number of transactions in a particular period.

(3) Monetary value of the purchase which refers to consumption money amount in a particular period.

RFM analysis is utilized in many ways by practitioners; therefore, RFM analysis can mean different things to different people. Classic RFM implementation ranks each customer on valuable parameters against all the other customers, and creates an RFM score for each customer/product. The first step is to sort the customer file according to how recently customers have purchased from the firm. Then database divided into equal quintiles and these quintiles are assigned the numbers 5 to 1. Therefore, the 20% of the customers who most recently purchased from the company are assigned the number
5; the next 20% are assigned the number 4, and so on. The next step involves sorting the frequently, monetary. Therefore, the database is divided into 125 roughly equal groups (cells) according to regency, frequency, and monetary value. Customers/products with high scores are usually the most profitable

LITERATURE REVIEW

E-business customer service by Anton, J., Hock, M.: e-Business Customer Service. The Anton Press, Santa Monica, CA. They defined CRM as a comprehensive business and marketing strategy that integrates technology, process, and all business activities around the customer. The widespread use of the web technology presents an opportunity for businesses to use the Internet as a tool for electronic customer relationship management (e-CRM). Gaps identified in this paper is that testing by means of the statistical analysis method of Structure Equation Model (SEM) yields results which are not applicable in real life scenarios as per our findings.[1]

E-CRM web service attributes as determinants of customer satisfaction with retail web sites by Richard Feinberg and Rajesh Kadam of International Journal of Service Industry Management. Despite widespread agreement that CRM/e-CRM has direct and/or indirect impact on customer satisfaction, sales, profit, and loyalty, the significance of e-CRM and the various e-CRM features in influencing customer satisfaction has not been well researched. According to Feinberg and Kadam, profits increase by 25–80% when customer retention rates increase by five points. Gaps identified in this paper are unable to determine both empirical and conceptual relationship between e-CRM features and customer satisfaction, and e- CRM features and customer loyalty.[2]

Managing Customers Relationships in the e-Business Economy by Jayanta Chatterjee of Journal of Scientific & Industrial research.

CRM as a discipline which focuses on automating and improving the business processes associated with managing customer relationships in the area of sales, management, customer service, and support. Gaps identified in this paper are unable to define beyond the need to create a more entrepreneurial atmosphere to cope with uncertainty and rapid market/consumer change.[3]

Customer relationship management

Since the early 1980s, the concept of customer relationship management in marketing, and consists of four dimensions: customer identification, customer attraction, customer retention and customer development has gained its importance. It is difficult to find out a totally approved definition of CRM. We can describe it as a comprehensive strategy and process of acquiring, retaining and partnering with selective customers to create superior value for the Company and the customer (Parvatiyar & Sheth, 2004). CRM is a comprehensive business and marketing strategy that integrates technology, process, and all business activities around the customer (Anton, 1996; Anton & Hoeck, 2002). Brown points out that CRM as ‘‘the key competitive strategy you need to stay focused on the needs of your customers and to integrate a customer facing approach throughout your organization’’ (Brown, 2000). Chatterjee also defines CRM as a discipline which focuses on automating and improving the business processes associated with managing customer relationships in the area of sales, management, customer service, and support (Chatterjee, 2000). According to Feinberg and Kadam, profits increase by 25–80% when customer Retention rates increase by five points (Feinberg & Kadam, 2002). CRM projects often fail and only about 40% of CRM implementations are successful (Feinberg & Trotter, 2001).

Customer loyalty

Creating a loyal B2B customer base is not only about maintaining numbers of customer overtime, but it is creating the relationship with business customers to encourage their future purchase and level of advocacy. Equipped with the knowledge of their business customers’ loyalty levels, a supplier will be able to figure how their endeavors to maintain good relationships can contribute to its profit levels. Some authors believe that loyal customers offer a steady stream of revenue for a company by remaining with the brand/supplier and rejecting the overtures of competitors (Lam, Shankar, Erramilli, & Murthy, 2004; Reich held & Teal, 1996). Considering this with the nature of large purchase and transactions in a B2B setting; there are gigantic rewards for those suppliers who succeed in creating and maintaining loyal customers. Some Authors have proposed several theories to link variables that one usually finds in relationship marketing and business marketing to the loyalty construct. In the B2B context, evidence shows that relationship elements affect customer loyalty. For example, Ricard and Perrien found that relationship practices have a direct impact on customer loyalty (Ricard & Perrien, 1999). Other Authors provide empirical evidence linking several constructs such as relationship quality, trust, involvement, satisfaction, purchase development, organizational change, and switching costs to influence B2B customer loyalty and retention (Chow & Holden, 1997; Eriksson & Vaghult, 2000). In some researches, customer lifetime value (CLV) or loyalty is evaluated in terms of recency, frequency, monetary

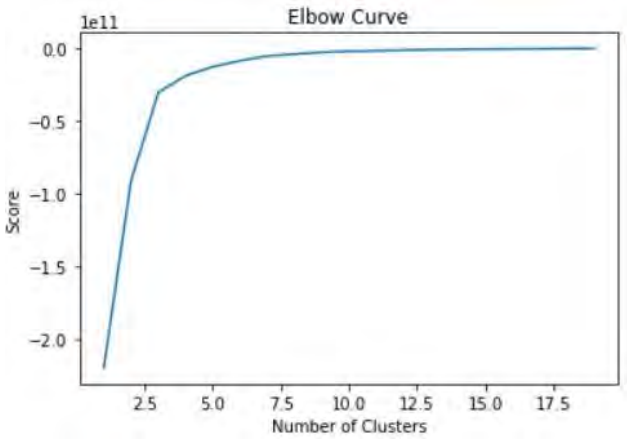
variables namely the integrating rate of each cluster, that is

Cj1 = wRCjR + wFCjF + wMCjM

where wR, wF, wM are the relative importance of the RFM variables (Liu & Shih, 2005). Other researches for classifying customer value have proposed a model based on computing the distance between the center of cluster and zero point as high value refers to most customer loyalty(Cheng & Chen, 2008).

METHODOLOGY

Steps for Customer Segmentation based on Loyalty:
1. Pre-processing of data by removing null values from the dataset and removing negative values from appropriate columns.
2. Transformation of data into the format required for RFM analysis. The dataset should be in the form of (Customer_ID, Recency, Frequency, Monetary)
3. Finding the optimal value of k using the elbow method.
The elbow method is a heuristic method of interpretation and validation of consistency within cluster analysis designed to help find the appropriate number of clusters in a dataset. It is often ambiguous and not very reliable, and hence other approaches for determining the number of clusters such as the silhouette method are preferable.



4. Applying k-means algorithm to cluster the data into segments based on customer loyalty.
Clustering is the process of grouping a set of physical objects into similar groups. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters (Han & Kamber, 2001). K
one of the well-known algorithms for clustering which is very sensitive to the choice of a starting point for partitioning the items into K initial clusters. We can compare the performance of different clustering methods using intraclass method when the grating mer n K-Means is number of fixed cluster of K value is defined as

$$F(K) = 1 \Big/ K\sum_{n=1}^{k}$$

Steps for Market Basket Analysis:
1. Pre-processing of data by removing null values from the dataset and removing negative values from appropriate columns.
2. Transform the data into format required for Apriori Algorithm.
3. Applying Apriori algorithm to mine frequent datasets
Apriori is an algorithm for frequent item set mining and association rule learning over relational databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.
Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.
Apriori uses breadth-first search and a Hash tree structure to count candidate item sets efficiently. It generates candidate item sets of length
sets of length k − 1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k - length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.
The pseudo code for the algorithm is given below for a transaction database T, and a support threshold of
Usual set theoretic notation is em
that T is a multi set. C k is the candidate set for level k. At each step, the algorithm is assumed to generate the candidate sets from the large item sets of the preceding level, heeding the downward closure

lemma. c o u n t [c] accesses a field of the data structure that represents candidate set
initially assumed to be zero. Many details are omitted below, usually the most important part of the implementation is the data structure used for storing the candidate sets, and counting their frequencies.

```
Apriori(T, ε)
    L_1 ← {large 1 − itemsets}
    k ← 2
    while L_{k−1} ≠ ∅
        C_k ← {c = a ∪ {b} | a ∈ L_{k−1} ∧ b ∉ a, {s ⊆ c | |s| = k − 1} ⊆ L_{k−1}}
        for transactions t ∈ T
            D_t ← {c ∈ C_k | c ⊆ t}
            for candidates c ∈ D_t
                count[c] ← count[c] + 1
        L_k ← {c ∈ C_k | count[c] ≥ ε}
        k ← k + 1
    return ⋃_k L_k
```

Conclusion

The proposed procedure has shown that for the purpose of clustering, when we combine the expanded WRFM model into K-means algorithm, we can see a tremendous improvement in classifying accuracy in order to reach to an excellent CRM. As the distance and the integrated rate of each cluster commonly used by many researchers as a separate and independent parameters, in this study the combination of these two parameters has been considered in clustering and classification analysis. The result of statistical test for model validation has shown that the developed methodology for CRM has an acceptable result with a high level of confidence in comparing with other commonly used models by researchers. The proposed CRM methodology can be used by industries as well as service sectors in evaluating their customers loyalty in a most efficient and effective manners. The statistical interpretation of how support, confidence and lift can correlate with marketing strategies take some time and know how on the field where it's being applied.

References

[1] ”Anton, J., & Hoeck, M. (2002). E-business customer service. Santa Monica, CA:

The Anton Press.

[2] “Brown, H. L. (2000). ‘‘The three Ws of technology”, customer relationship management: A strategic imperative in the world of e-business (pp.

219–225). Toronto: Wiley

[3] “Chatterjee, J. (2000). Managing customer relationships in the e-business economy.

Journal of Scientific and Industrial Research, 749–
752.

[4] “Feinberg, R., & Kadam, R. (2002). E-CRM web service attributes as determinants of

customer satisfaction with retail web sites. International Journal of Service

Industry Management, 13(5), 432–451.

[5]”Feinberg, R., & Trotter, M. (2001). Immaculate deception: The unintended negative

effects of the CRM revolution: May be we would be better off without customer

relations management. In D. H. Renner (Ed.), Defying the limits, Vol. 2,

Montgomery Research, San Franscisco, CA

[6] ”Shina, H. W., & Sohnb, S. Y. (2004). Segmentation of stock trading customers

according to potential value. Expert Systems with Applications, 27–33.

[7] “Cheng, C.-H., & Chen, Y.-S. (2008). Classifying the segmentation of customer value

via RFM model and RS theory. Expert Systems with Applications.

[8] “Eriksson, K., & Vaghult, L. A. (2000). Customer retention, purchasing behavior and

relationship substance in professional services. Industrial Marketing

Management, 29, 363–372

[9] “Han, J., & Kamber, M. (2001). Data mining: Concepts and techniques. San Francisco:

Morgan Kaufmann Publishers.

[10] “Parvatiyar, A., & Sheth, J. (2004). Conceptual framework of customer relationship

management. In J. N. Sheth, A. Parvatiyar, & G. Shainesh (Eds.), Customer

relationship management: Emerging concepts, tools and applications (5th ed.).

New Delhi: Tata McGraw-Hill Publishing.

Analyzing Vocal Patterns to Determine Emotions

Dr. Rekha Sharma, Abhinav Birajdar, Manas Raina, Mehul Suthar
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
abhinavbirajdar18@gmail.com, raina.manas@gmail.com, mehulsuthar1999@gmail.com

Abstract— **The human voice is very versatile and carries a multitude of emotions. Emotion in speech provides additional insights about human actions. Through analysis of those practical features of speech, we can better understand people's intentions, whether they are unhappy customers or fans. Man is able to easily determine a speaker's emotion, but the field of emotion recognition through machine learning is an open research area. We begin to study emotion in speech by finding an emotion. In particular, we examine the classification of anger in speech samples. In our analysis of emotions, we begin by delineating the data used, and through this analysis, we investigated whether CNN is the best algorithm for selecting features that are relevant for predicting emotions. [8]. We also consider several machine learning models with which emotions are classified. A variety of machine learning techniques for the analysis of vocal patterns to determine emotions have been proposed with varying levels of success and accuracy. However, the accuracy of each technique is based on the number of machines under consideration and the machine learning tools / techniques used [1] [5]. The accuracy of emotion is high when using a large number of attributes and examples.**

Keywords—Delineating, CNN (Convolutional Neural Network)

I. INTRODUCTION

The human voice consists of sounds that are used by a person to use acoustic folds, such as to perform acoustic activities such as talking, singing, laughing, shouting, etc. Frequency of the human voice is particularly a part of the human sound generation system with vocal cords or folds being the primary source of the sounds produced. Other sound production mechanisms originating from the same general area of the body include the production of inconsistent consonants, clicks, whistles, and whispers.

Generally, the mechanism for generating human voice can be divided into three parts; Lungs, larynx folds and articulators The human voice and associated speech patterns can be characterized by several features, the primary ones being pitch, loudness or sound pressure, timbre and tone [2]. Pitch is an auditory sensation in which a listener assigns musical tones to relative positions on a musical scale based primarily on their perception about the frequency of vibration [5] [7]. Pitch can be quantified as a frequency, but is based on the subjective perception of a sound wave. Sound oscillations can be measured to obtain a frequency in hertz or cycles per second. Pitch is independent of the intensity or amplitude of the sound wave. A high-pitched sound indicates fast oscillations, whereas, a low-pitched sound corresponds to slow oscillations. The pitch of complex sounds such as speech and music notes corresponds to the repetition rate of periodic or nearly-periodic sounds, or the reciprocal of time intervals between similar repetitive events in the sound wave. Loudness is a subjective perception of sound pressure and can be defined as a characteristic of auditory sensation, in the context of which, sounds can be ordered on a scale ranging from quiet to loud [8]. Sound pressure is the local pressure deviation from ambient, average or equilibrium atmospheric pressure, which is caused by the sound wave [9]. Sound pressure level (SPL) is a logarithmic measure of the effective pressure of a sound relative to a reference value and is often measured in units of decibels (dB). The lower limit of hearing is defined as SPL of 0 dB, but the upper limit is not clearly defined. Timbre is the perceived sound quality of a musical note, sound, or tone [5]. Timbre distinguishes different types of sound production and enables listeners to distinguish different instruments in the same category. Physical characteristics of sound that determine the perception of time include spectrum and envelope. Figure 1 shows a sound wave. In simple words, timbre is that one particular sound is considered different from another sound, even if they are the same pitch and louder.

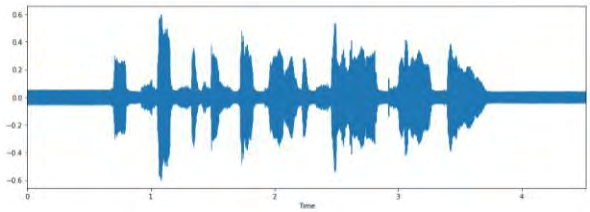


Figure 1: Waveform of sample audio file

II. PROPOSED WORK

This section describes an algorithmic approach to devoting human emotions through voice and speech-pattern analysis. To achieve this objective, three test cases were investigated, corresponding to three emotional states: normal emotional state, angry emotional state, and nervous emotional state. To carry out the analysis, four vocal parameters will be taken into consideration: pitch, SPL, timbre, and time interval between consecutive words of speech. To represent a time table quantitatively, its temporal envelope for advance and decay times will be considered. The primary function of the proposed algorithmic approach is to explore the different emotional states by analyzing deviations from the general emotional state to the above four parameters. Pitch of sound is the characteristic of sound by which an acute note can be distinguished from a grave or flat note. If a pitch is high, the sound is called sharp and if the pitch is low, the sound is flat. The pitch of sound is determined by the rate of vibration or frequency of the sound wave. Pitch is simply the rate at which vibrations are produced. It is usually expressed as the number of hertz. Sound pressure level (SPL) is a logarithmic measure of the effective pressure of a sound relative to a reference value and is often measured in units of decibels (dB). Timbre is the perceived sound quality of a musical note, sound, or tone. Timbre distinguishes different types of sound production and enables listeners to distinguish different instruments in the same category. Physical characteristics of sound that determine the perception of time include spectrum and envelope. The time interval between words is the average difference of time between two consecutive words in speech.

• Case 1: General Emotional Status This test case consists of pitch, SPL, timbre, and word-time interval data obtained from speech samples that were at the speaker's comfort and general emotional state. This test case serves as the basis for the remaining two test cases. All parameter statistics represent the average values obtained from the indicated samples. As shown in Table I, for the purpose of demonstration, data from two speech samples have been analyzed.

	Pitch (Hz)	SPL (dB)	Timbre ascend time (s)	Timbre descend time (s)	Time gaps between words (s)
Speech Sample 1	1248 Hz	Gain -50 dB	0.12 s	0.11 s	0.12 s
Speech Sample 2	1355 Hz	Gain -48 dB	0.06 s	0.05 s	0.12 s

T able 1: Average values of vocal statistics obtained from recorded speech samples for a normal emotional state

Case 2: Angry Emotional State This test case contains pitch, SPL, timbre, and word-timing gap statistics derived from speech samples, which are characterized by increased pitch, usually in an emotional state stimulated by the speaker. . All parameter statistics represent mean values obtained from speech samples, as shown in Table II. The same speech samples that were previously used in Case 1 have been used in Case 2, but with a distinctly distinct form of an agitated or angry emotional state.

	Pitch (Hz)	SPL (dB)	Timbre ascend time (s)	Timbre descend time (s)	Time gaps between words (s)
Speech Sample 1	1541 Hz	Gain – 30 dB	0.13 s	0.10 s	0.09 s
Speech Sample 2	1652 Hz	Gain – 29 dB	0.06 s	0.04 s	0.10 s

Table 2: Average values of vocal statistics obtained from recorded speech samples for an angry emotional state

Case 3: stereotypical emotional state This test case consists of pitch, SPL, timbre, and word-time interval data obtained from speech samples that were in the speaker's nervous or overwhelmed emotional state. Speech samples that were previously used in Case 1 have been used in Case 3, but with a different intonation typical of a panicked emotional state, as shown in Table III.

	Pitch (Hz)	SPL (dB)	Timbre ascend time (s)	Timbre descend time (s)	Time gaps between words (s)
Speech Sample 1	1443 Hz	Gain – 46 dB	0.13 s	0.09 s	0.13 s
Speech Sample 2	1560 Hz	Gain – 44 dB	0.07 s	0.04 s	0.14 s

Table 3: Average values of vocal statistics obtained from recorded speech samples for a panicked emotional state

From the three table it can be seen that the average pitch for a speech sample of normal emotional state is 1300Hz as compared to 1500Hz for panicked emotional state and 1600Hz for angry emotional state. Hence pitch increases in going from normal to panicked to angry emotional state.

Similarly, for sound pressure level (SPL), angry emotional state has relatively lower SPL followed by panicked emotional state, with the normal emotional state having highest SPL. The time gap between words is the least in angry emotional state as a person in angry state speaks faster. Normal emotional state and panicked emotional state, on the other hand, have comparable time gaps between words

REQUIREMENTS
Software Requirements-

- Front End-Pyqt5
- Back End-ML Model using Python on Jupyter
- Operating System

Hardware Requirements-

- Min Memory: 4GB
- Storage size: 10GB
- Dedicated Graphics Memory
- Processor: Intel i3 or higher Training audio

III. ARCHITECTURE

Dataset:

This dataset consists of 1500 audio file inputs from 24 different actors. 12 men and 12 women where these artists record short audio in 5 different emotions, namely 1 = neutral, 2 = calm, 3 = happy, 4 = fearful, 5 = angry. Each audio file is named in such a way that the 7th character corresponds to the different emotions they represent.

Libraries to be used:

LibROSA:

Librosa is a python package for music and audio analysis. It provides the necessary building blocks to create music information retrieval systems. Librosa's load function will read in the path of an audio file and return a tuple with two items. The first item is an 'audio time series' (type: array) for the audio track. The second item in the tuple is the sample rate that was used to process the audio.

Keras:

Keras is a high-level neural network API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. This allows for easier and faster prototyping. It supports the combination of both a compatible network and a recreative network as well as both.

TensorFlow:

TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used directly to create a deep learning model, or by using a wrapper library that simplifies the process built on top of TensorFlow.

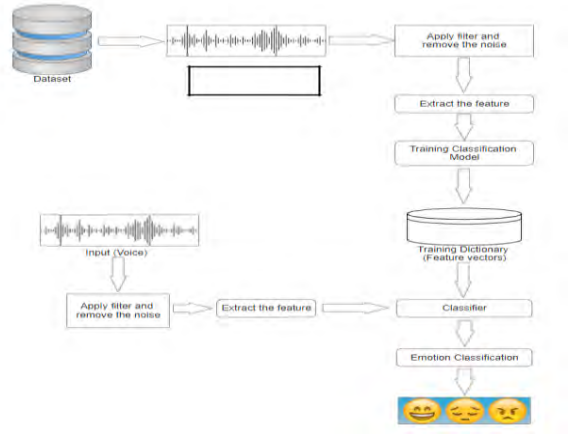


Figure 2: Architecture of model

IV. TECHNOLOGY USED

Front end: The following techniques will be used to design the structure of the project:

1) PyQt5:
PyQt5 is a comprehensive set of Python bindings for Qt v5. It is implemented as more than 35 extension modules and enables Python to be used as an alternative application development language for C ++ on all supported platforms including iOS and Android.

PyQt5 can also be embedded in C ++ based applications to allow users of those applications to configure or enhance the functionality of those applications.

Back End: The back end technique used in the project:
1) Python:

Python is an interpretive, high-level, general-purpose programming language. Created by Guido van Rosum and first released in 1991, Python's design philosophy emphasizes code readability with notable use of critical WhatsApp. Its language-building and object-oriented approach aims to help programmers write clear, logical code for small and large scale projects. Python is dynamically typed and garbage is collected. It supports several programming paradigms, including procedural, object-oriented and functional programming.
One of the aspects that makes Python in general such a popular option is its abundance of libraries and frameworks that facilitate coding and save development time. Machine learning and deep learning are exceptionally well crafted.

NumPy used for scientific computation, SciPy for advanced computation, and scikit-learn for data mining and data analysis are among the most popular libraries with heavy-hitting frameworks such as TensorFlow, CNTK, and Apache Spark Work. In terms of machine learning and deep learning, these libraries and frameworks are in essence Python-first, while some, such as PyTorch, are written specifically for Python.

2) Jupyter:

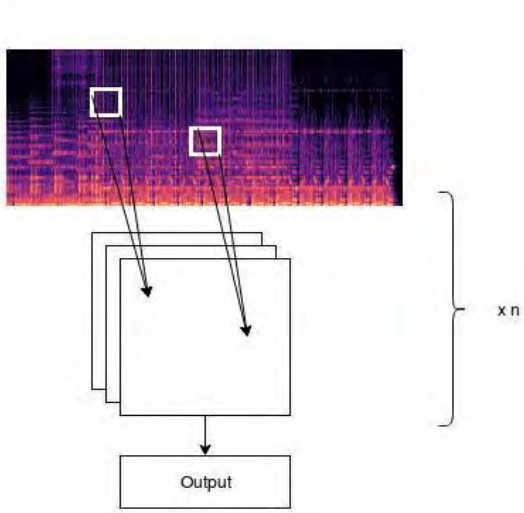
Jupiter Notebook (formerly iPython Notebook) is a web-based interactive computational environment for creating Jupiter Notebook documents. The term "notebook" can generally refer to several different entities, mainly Jupiter Web Application, Jupiter Python Web Server, or Jupiter Document Format depending on the context. A macro notebook document is a JSON document that has a version schema, and contains an ordered list of input / output cells that can contain code, text (using markdown), mathematics, plots, and rich media, typically ending with ".ipynb". Detailed.

Jupiter Notebook can connect to multiple kernels to allow programming in multiple languages. By default Jupiter notebook ships with the iPython kernel. As of the 2.3 release (October 2014), there are currently 49 Jupiter-compatible kernels for several programming languages, including Python, R, Julia, and Haskell.

The notebook interface was added to IPython in the 0.12 release (December 2011), renamed Jupyter Notebook in 2015 (IPython 4.0 - Jupyter 1.0).

V. METHODOLGY

The log-mail spectrogram is an advanced method for classifying audio using a mail-spectrogram instead of a raw audio wave. The mel spectrogram represents the short-term power spectrum of a sound, which is based on the linear cosine conversion of the log-in spectrometer spectrum. Frequency of a nonlinear mel scale. The mel spectrogram transforms the input raw sequence into a 2D feature map where one dimension represents time and the other represents frequency and the value represents amplitude. At a high level, each audio file from the repository is processed and



used to create a feature vector with a label corresponding to emotion.

Figure 3: Log-Mel Spectrogram

The next step involves removing features from audio files that will help our model learn between these audio files. For feature extraction we use the Librosa library in Python which is one of the libraries used for audio analysis. Here are some things to note. When removing features, all audio files are timed for 3 seconds to achieve the same features. The sampling rate of each file is doubled, keeping the sampling frequency constant to obtain more features, which will help classify the audio file when the size of the dataset is smaller. After extracting the features, we apply the feature selection algorithm to determine the most relevant results and apply different models to the features with the highest scores. The most straightforward technique involves determining the average energy of the signal. This metric indicates the "volume" of the speaker, along with the total energy

in the signal. Duration also provides insight into emotion, as do statistics such as maximum, minimum, range, mean, and standard deviation of both the signal and the spectrum. These may indicate fluctuations in volume or pitch that may be useful in determining emotions. We also process signals in the frequency domain via the (fast) Fourier transform. We use window samples to obtain an accurate representation of the frequency content of the signal at different points in time. Taking the square value of the signal in each window sample, we can obtain the power spectrum. We use the values of the power spectrum as features, but we also find frequencies that have the greatest power. We obtain the three largest frequency peaks for each window and combine them into a feature vector. In addition, we detect maximum and minimum frequencies with sufficient power for each time frame and use these values to determine the frequency range for each frame. The auditory spectrum can be obtained in a uniformly spaced interval by mapping the power spectrum to the auditory frequency axis by combining forest transformer bins.

VI. RESULT AND DISCUSSION

From what has been reviewed above, it can be said that the algorithm performs with an accuracy of 72%. The 5 emotions, 10 possibilities considering both male and female voices, can be classified successfully. The classification of male voices from female voices has been achieved with 100% accuracy. Above planned comparisons of sample feature table reveal that different features of an audio file have different ranges, responsible for varying emotions. For two speech samples of different emotions some attributes may be similar but they will never be identical. Results (Average Precision):

1 D: 0.754
2D: 0.849
Average predictions of both: 0.883

The 2D mail model outperforms the 1D raw wave model but the average of the two different models is quite different. This is probably because each model learns different representations and makes different types of mistakes and on average each model fixes the errors of the other in some way.

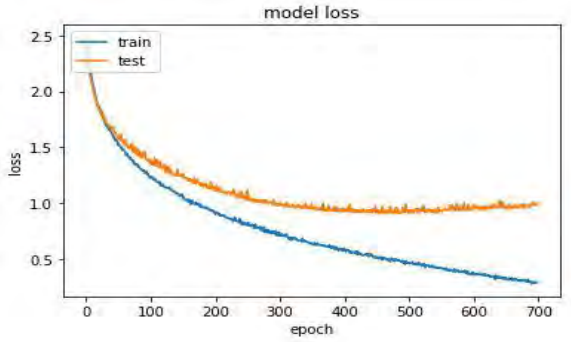


Figure 4: Graph depicting model loss

As seen in the figure above, the loss in testing set for every successive epoch decreases at a lesser rate than the training dataset resulting in an overall lesser loss in the testing phase.

VII. CONCLUSION

Model construction was a challenging task as it involved a lot of trial and error methods, tuning and so on. The model is very well trained to distinguish between male and female voices and is isolated with 100% accuracy. The model was designed to detect emotions with over 70% accuracy. Accuracy can be increased by including more audio files for training. Reducing human emotion through voice and speech analysis is a practical feasibility and may possibly be beneficial for improving human conversational and persuasion skills. This paper presents an algorithmic approach to the detection and analysis of human emotions based on voice and speech processing. Three test cases have been investigated, which relate to three emotional states: general emotional state, angry emotional state, and nervous emotional state. Each case displays vocal features associated with the trait that may help differentiate the associated emotional state. As the scope of future work, more complex emotional states can be analyzed such as cheerful, sad, brooding, surprise, etc. When conducting the proposed study, the speech data were sampled at a rate of ten sample points per second. To improve the accuracy of the mean attribute values, a greater number of sample points can be considered. Furthermore, in the proposed approach, only four characteristics were analyzed to detect different emotional conditions. To improve recognition accuracy, other secondary vocal features may also be evaluated.

References

[1] Maisy Wieman, Andy Sun, Analyzing Vocal Patterns to Determine Emotion
[2] Poorna Banerjee Dasgupta,Detection and Analysis of Human Emotions through Voice and Speech Pattern Processing, IJCTT, Volume 52 Number 1, October 2017
[3] Mitesh Puthran, Speech-Emotion-Analyzer
[4] K. H. Kim, S. Bang, S. Kim, Emotion recognition system using short-term monitoring of physiological signals, Medical and biological engineering and computing 42 (3) (2004) 419–427.
[5] Alice V. Hatfield, and Danielle Sova, Machine Learning Analysis of Human Emotion from Speech Signals, George Mason University, December, 2015
[6] Assel Davletcharova, Sherin Sugathan, Bibia Abraham, Alex Pappachen Jamesa, Detection and Analysis of Emotion From Speech Signals, Procedia Computer Science , 2015.
[7] www.towardsdatascience.com
[8] www.kaggle.com
[9] www.keras.io

Design and Development of Agricultural Drone for Plant Health Analysis

Mrs. Vaishali Nirgude, Yogesh Bharambe, Divesh Singh, Vedang Binsale
Computer Department
Thakur College of Engineering and Technology
Mumbai, India
vaishali.nirgude@thakureducation.org, yogesh01bharambe@gmail.com, diveshsingh2210@gmail.com, vedang.binsale@gmail.com

Abstract— **Agricultural drones let farmers see their fields from the sky. This overview of the field can reveal many issues such as irrigation problems, soil variation, and pest and fungal infections. The combination of these factors allows the farmer to distinguish between healthy and unhealthy plants, a difference not always clearly visible to the naked eye. Thus, these views can aid in assessing crop growth and production. The project is to develop an autonomous flying quadcopter, fitted with a GPS tracking system and programmed to be able to autonomously fly from one location to a different location using GPS coordinates. Its goal is to act as a proof-of-concept for a small scale autonomous geographical area mapping and find plant health using VARI algorithm.**

Keywords— VARI (Visible Atmospherically Resistant Index), NDVI (Normalized Difference Vegetation Index), UAV (Unmanned Ariel Vehicles), GIS (Geographical Information System), GPS Module and Telemetry

I. INTRODUCTION

Remote sensing supported drones (UAV- Unmanned Aerial Vehicle) is a rapidly developing field of technology, which allows new possibilities for monitoring of the environment. Due to technological innovations lightweight multi- and hyper spectral sensors have become available which may be carried by small drones. The applications for UAV can comprise precision agriculture, vegetation monitoring and forest monitoring [1]. Drone based mapping enables to map agricultural lands with a very high spatial resolution, which enables to observe even small anomalies in vegetation [2]. Such anomalies in vegetation/ plants could be caused by, for example, drought induced uneven germination and growth, seedling loss caused by flooding and anoxia, disease outbreak, pest invasion, unbalanced nutrient availability and uptake, and overwintering damages [2] [3] [4].

II. LITERATURE SURVEY

Nandyal Suvarna - "Agridrone: Automation of Agriculture using IoT". IJISRT Journal (2019). This paper discusses a system which consists of a flying drone fitted with a camera eye to record images of crops in a scheduled time. The work involves developing an intelligent system by building a knowledge-base to guide agriculturists. The knowledge-base includes various cases of various crops and decisions based on crop image analysis. From captured images, the parameters that are planned for analysis by image processing are The image of the plant will be acquired using the
external camera eye fitted on drone which is interfaced with Raspberry Pi along with other different sensors modules through IOT. This image is then diagnosed to measure moisture content in soil and amount of green in leaves.

Ahmad - "Aerial mapping using high resolution digital camera and unmanned aerial vehicle for Geographical Information System." 2010 6th International Colloquium on Signal Processing & its Applications, IEEE (2010). This study concentrates on the use of high-resolution digital camera and a very light platform known as unmanned aerial vehicle (UAV) as data acquisition system in capturing digital aerial photographs. The acquired digital aerial photographs were processed using image processing software to produce digital map and digital orthophoto. The results showed that an accuracy of sub-meter can be obtained using the employed method. In Geographical Information System (GIS), it is quite common that topographic map and orthophoto are used as a base map. Hence, the findings from this study could also be used as an input for GIS. Reinecke - "The influence of drone monitoring on crop health and harvest size." 2017 1st International Conference on Next Generation Computing Applications (NextComp), IEEE (2017). This paper looks at the benefits of drones in agriculture, and their limitations, illustrating from examples how drones operate on farms. Different features of drones are discussed, specifically how they assist farmers in maximizing their harvest by detecting problems early

and managing the crops by using specific cameras to detect pests and water shortages.

Santos - "Indoor waypoint UAV navigation using a RGB-D system." 2015 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS), IEEE (2015). This paper presents a method for estimating the orientation and position of an unmanned aerial vehicle (UAV), through processing information provided by a RGB- D sensor and an inertial measurement unit (IMU) in order to inspect predefined points in an indoor environment.

Roberts - "GPS Guided Autonomous Drone.", University of Evansville (2016). The project is to design an autonomous flying drone, specifically a quadcopter. The drone is fittedwith a GPS tracking system and programmed to be able to autonomously fly from one location to another using GPS coordinates. The goal of the project is to act as a proof of concept for small scale autonomous aerial delivery similar to that nearing deployment by Amazon.

F.G. Costa and et.al - The use of unmanned aerial vehicles and wireless sensor network in agriculture applications. IEEE Int. Geoscience and Remote Sensing Symposium, 2012. Unmanned aerial vehicles (UAVs) represent technological developments used for precision agriculture. They provide high-resolution images taken from crops and when specific indices are applied, useful outputs for farm management decision-making are produced. The current paper provides a literature review on the use of UAVs in agriculture and specific applications are presented

III. EASE OF USE

A. Monitoring Crop Production

Drones have already proven to be effective at recording the canopy reflectance over vineyards. To measure the growth of crops over a season, aerial images can be taken at the start of the season, at predetermined intervals during the season and just prior to harvesting to illustrate the growth across the field, as well as highlighting any rows that show signs of stinted growth due to poor irrigation or low initial nitrogen content [3]. Birds may fear a scarecrow for a time, but they will soon realize it is not alive and return in numbers –meaning the only solution is a lethal one. Larger animals like baboons and deer can only be kept away with tall often electrified fences with empty cans to make a noise. In the midst of all these preventative measures, farmers may forget about a section of an orchard or field – the consequences of which will only be apparent when it is too late to remedy. Crop diseases can be devastating and classified as fungal, bacterial or viral. Drones equipped with Infrared cameras can see inside plants, giving a clear image of the condition thereof. If a farmer can detect an infection before it spreads, preventative measures can be taken

Like removing the plant -before the infection spreads to neighboring plants. Drones can be equipped with a multispectral camera that can detect the water content underground, which can allow a farmer to determine if a crop row is parched or over hydrated. Farmers can adjust their irrigation until even water table is created, eliminating water as a source of crop distress. If multispectral camera is not available, the drone can still take overhead pictures of the fields and highlight areas where plants seem discolored or smaller than the rest of the field. This technique can also be used to detect underground leaks in existing irrigation systems [2] [4].

B. CRIME AND DAMAGE REPORTS

The drone can aid the authorities in identifying the vehicle employed by the criminals, speeding up the investigation. Also, some farmers have employed drones as guards to patrol entrances and along fences.

C. Forestry

Open Forest can use drone-based forest and landscape mapping to provide a new perspective for valuation, monitoring and research. Hundreds of pictures taken by drones are mapped together to large and high resolution orthomaps. These orthomaps can then be used for analysis, planning and management via the GIS systems [2]. Nova drone uses drone technology to improve forest management and operational planning, including the monitoring of illegal activities.

IV. PROPOSED WORK

For this Project, we have built a Quadcopter Drone for agriculture surveillance and analysis. The additional functionality/component is "*ublox neo 7m gps with compass*"

– it is a gps component with inbuilt compass in it; it will be used to guide the drone without any manual guidance from any personnel from organization [4] [5].

With the help of Telemetry components, the directions to the drone can be given from any android phones through an application. Another component is "*Xmate 16 Mega pixel Waterproof Camera*" – Used for recording the whole agricultural land and take HD Images for further analysis.

After Flight and recording of data of crop, it will be passed through a set of code. We will use VARI (Vegetation Atmospheric Resistant Index) Algorithm for Analysis of Crop health.

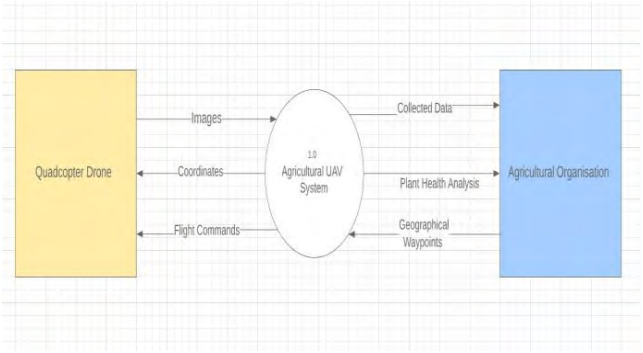


Figure 1: Structural model - DFD Level 0

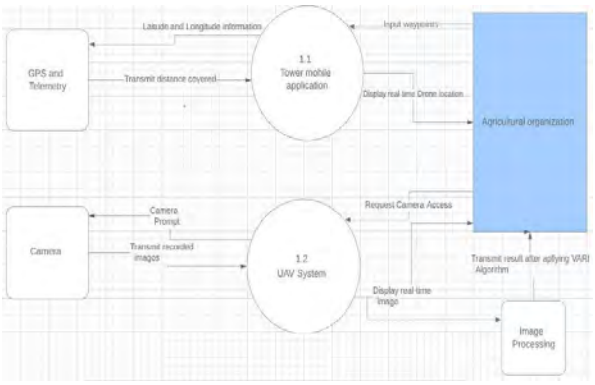


Figure 2: Structural model - DFD Level 1

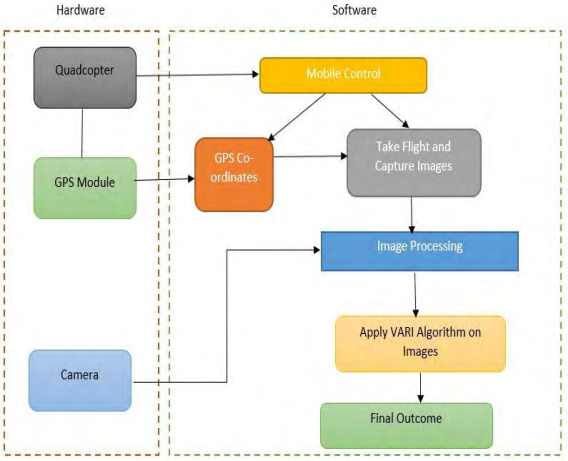


Figure 3: Flow Diagram

A. VARI Algorithm

VARI Index can be given as:

$$VARI = \frac{Green - Red}{Green + Red - Blue}$$

VARI algorithm is a vegetation index used to record plants health. For VARI algorithm, RGB (standard) cameras, which capture Red, Green, and Blue light, is used for imaging purposes. The Visible Atmospherically Resistant Index (VARI) was designed to work with RGB cameras. VARI is a measure of "how green" a picture is. VARI is meaningful when working with non-NDVI imagery, hence it is not intended as a substitute for NDVI algorithm. RGB images with the VARI algorithm applied make it possible to detect areas of crop stress in a vegetation field [5] [7].We can measure plant health using drone imagery because healthy plants

reflect light differently than unhealthy plants. Plants that are healthier tend to reflect more green light than red light, which is why they appear green. Plants also reflect near-infrared light that's invisible to the naked eye but are often detected with near-infrared sensors.

Plant health algorithms like VARI and NDVI use the data generated by the proportions of sunlight captured across different bands (red, green, blue and sometimes near-infrared) to compute numerical values (between 0 to 1) for every pixel of a given drone image/ map. Maps with plant health algorithms are then assigned color based on those numerical values, which makes it easy to spot unavoidable plant diseases [6].

B. GPS MODULE AND TELEMETRY

A GPS module can provide an estimation of your drone's altitude. However, GPS modules are rather inaccurate and can only offer you a position to within 5m. However, by combining measurements for other sensors, the flight controller can get a better picture of what the drone is doing. The main feature employed by the GPS module is that you simply can autonomously fly your drone to way-points, so your drone can potentially fly on its own from takeoff to landing.

The telemetry modules are the actual radio devices, which aids in waypoint navigation.

C. APPLICATIONS

- Land surveying
- Cadastral maps
- Topographical surveys

Precise measurements

- Land management and development
- Monitor plant health
- Perform plant count
- Optimize plant ROI

D. ADVANTAGES

Easy to use

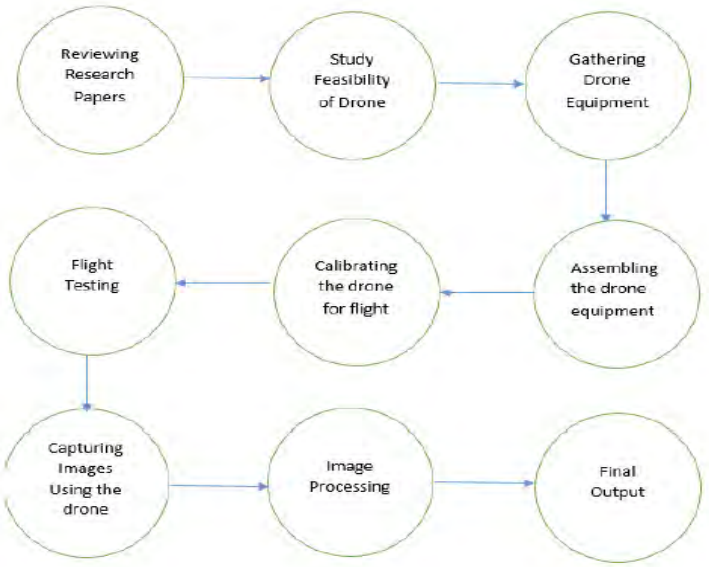
Cost effective

- Automated; requires less human interaction due to GPS and Telemetry
- Can be controlled through any Android mobile
- Recorded images pass through VARI algorithm, yielding plant's health

E. METHODOLOGY

For this project, Scrum methodology is used which is an agile software development methodology. By adopting this methodology, every team member can know the progress of the project. Following this model, our method can be described by the following flow graph:

Figure 4: Methodology



F. Drone Connections

Figure 5: Flight Controller Connections

The above Figure shows the connection of all components to APM 2.8 processor for automated operation.

V. RESULTS AND DISCUSSION

An autonomous drone such as the one constructed in this project (a Quadcopter) can have many uses in practical applications. Among other things, a GPS guided drone could be used for military and defence purposes.

Hence, there is no doubt that the combination of the UAV and the digital camera could be used and explored further for GIS application, mapping and other applications. However, user must be aware that the area covered by the digital camera is very small.

The major advantage of Drone is the auto landing capability which reduces the risk factors designed in a simple and cost-effective manner. It also helps in analysing plant quality and health by using VARI algorithm. Regarding the functioning of the drone, we have used Mission Planner application to program the drone to take flight. For formation of orthomosaic maps and calculating VARI indexes, we would be using WebODM application.

The top view of final constructed quadcopter is:

Figure 6: Top View of Quadcopter

To determine plant health, our quadcopter captures images of the fields/ area and processes them via VARI algorithm. The program outputs a value between 0 to 1; 0

Indicating bad plant health and 1 indicating good plant health.

The sample output (health) for good and bad crops, processed by VARI algorithm are:

Figure 7: VARI Index for bad crops

Figure 8: VARI Index for good crops/ plants

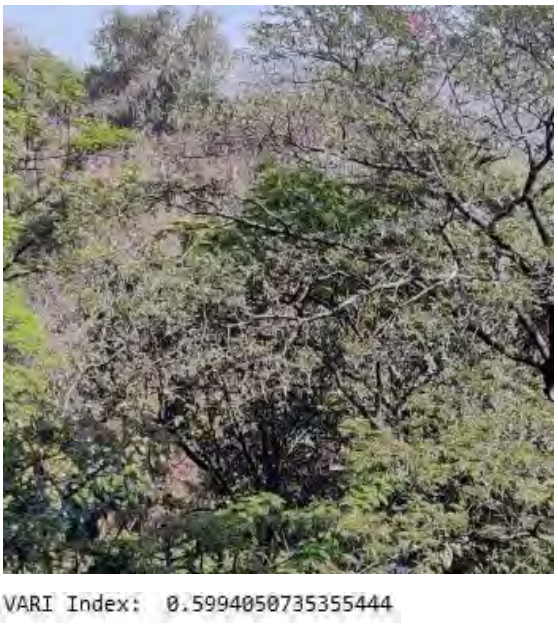


Figure 8: VARI Index for good crops/ plants

ACKNOWLEDGMENT

We wish to express sincere thanks to our guide Mrs. Vaishali Nirgude and our Computer department for providing research opportunity, so that we can research all this about Agricultural drone. We also extend our heartfelt thanks to our colleagues, family members and well- wishers.

References

[1] Nandyal, Suvarna, Priya S. Khamitkar and Priyanka S. Joshi. “Agridrone: Automation of Agriculture using IoT.”, *IJISRT Journal (2019)*

[2] Ahmad, Anuar, and Abd Manan Samad. “Aerial mapping using high resolution digital camera and unmanned aerial vehicle for Geographical Information System.” *2010 6th International Colloquium on Signal Processing & its Applications, IEEE (2010)*.

[3] A. Ahmad. “Mapping using small format digital imagery and unmanned aerial platform”. In *South East Asia & Survey Congress, 2009*, 4-6 August; Bali, Indonesia

[4] Reinecke, Marthinus, and Tania Prinsloo. “The influence of drone monitoring on crop health and harvest size.” *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, *IEEE (2017)*.

[5] Gitelson, Anatoly A., et al. “Novel algorithms for remote estimation of vegetation fraction.” *Remote sensing of Environment* 80.1 (2002): 76-87.

[6] Ortiz, B. V., et al. “Determination of differences in crop injury from aerial application of glyphosate using vegetation indices.” *Computers and electronics in agriculture* 77.2 (2011): 204-213.

[7] Santos, Milton CP, Mário Sarcinelli-Filho, and Ricardo Carelli. “Indoor waypoint UAV navigation using a RGB-D system.” *2015 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS)*, *IEEE (2015)*.

[8] Roberts and Cameron. “GPS Guided Autonomous Drone.”, *University of Evansville* (2016).

[9] F.G. Costa and et.al, “The use of unmanned aerial vehicles and wireless sensor network in agriculture applications”, *IEEE Int. Geoscience and Remote Sensing Symposium, 2012*.

[10] J. Primicerio et al -, “A flexible unmanned aerial vehicle for precision agriculture.” *Precision Agriculture 13(4), 517-523, 2012(Volume:115, Issue:7)*

[11] E. George, G. Tiwari, R. Yadav, S. Sadana, and E. Peters, “Uav systems for parameter identification in agriculture,” *in Proc. GHTCSAS, 2013, pp. 270–273*.

[12] Zhou, X., et al. “Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery.” *ISPRS Journal of Photogrammetry and Remote Sensing* 130 (2017): 246-255.

[13] Identifying Crop Variability Patterns. [Online]. Available: https://blog.dronedeploy.com/identifying-crop-variability- whats-the- difference-between-ndvi-false-ndvi-and-vari- plant-health- 98c380381a33

[14] WebODM. [Online]. Available: https://github.com/OpenDroneMap/WebODM

[15] Vegetation Indexes. [Online]. Available: http://web.pdx.edu/~nauna/resources/8-2012_lecture1-vegetationindicies.pdf

Personalized Treatment Based On Patient History

Rohit Sharma, Nitesh Tiwari, Ganesh Wartale
Department of Computer Engineering
Thakur college of Engineering and technology
Mumbai, India
rohit.tps123@gmail.com, niteshtiwari1971@gmail.com, wartaleganesh@gmail.com

Abstract— **Health care practices involve collecting all kinds of patient data which would help the doctor correctly diagnose the condition the subject is likely suffering from. This data could be everything from the simple symptoms observed by the subject, initial diagnosis by a physician or a detailed test result from a lab. Thus far this data is only utilized for subjective analysis by a doctor who then ascertains the disease in play using his/her personal medical expertise. We posit that there is definite potential for application of data mining routines on this rich reserve of patient data. Employing apposite data mining techniques, useful patterns and conclusions could be drawn from the raw data at our disposal. These findings could in turn be utilized in a number of productive ways like to carry out automated diagnosis, equip doctors with a better understanding of the causes and factors in play behind a particular disease. Automated diagnosis in particular could prove very useful for the determination of non-critical diseases or in cases where a doctor may not be available to carry out diagnosis such as in case of being in a remote location.. We also aim to incorporate it in an application which will provide diagnosis, storing of patient information and provide visualization of that data.**

I. INTRODUCTION

Health care practices involve collecting all kinds of patient data which would help the doctor correctly diagnose the condition the subject is likely suffering from. This data could be everything from the simple symptoms observed by the subject, initial diagnosis by a physician or a detailed test result from a lab. Thus far this data is only utilized for subjective analysis by a doctor who then ascertains the disease in play using his/her personal medical expertise. We posit that there is definite potential for application of data mining routines on this rich reserve of patient data. Type 1 diabetes can occur in childhood or adolescence age. Type 2 diabetes usually affects the adults who are obese. In this type, the body resists observing insulin or fails to produce insulin. Type 2 generally occurs in the middle or aged groups. Moreover, there are other causes for diabetes such as bacterial or viral infection, toxic or chemical contents in food, autoimmune reaction, obesity, bad diet, change of lifestyles, eating habit, environment pollution, etc. Diabetes leads to various diseases such as cardiovascular complications, renal issues, retinopathy, foot ulcers, etc.

A. Problem Definition

Going to a doctor isn't always possible, there are plenty of reasons for this, the area could be facing with an epidemic causing the doctors to be occupied with the immediate case, the person could fall ill in the middle of the night, the problem could be severe and might require some information about the problem itself and how to make it bearable until the patient is taken to a doctor. It becomes necessary to provide an automated diagnosis tool to the people in order to address these issues.

II. LITERATURE REVIEW

Following is some of the search which has been reviewed for the proposed system: Shadegg Behave proposed, this system that comes under the category of data mining. The system performs data mining on patterns and correlation to predict the economic events. This system utilizes K-Nearest Neighbor for estimating values that will maintain a strategic distance from financial distress and bankruptcy [8]. In the current review k- Nearest Neighbor characterization technique, have been examined for economic estimating. Lately, after the situation of worldwide financial emergency, the quantity of bankrupt organizations has risen. Since organizations' financial distress is the principal phase of bankruptcy,[3] utilizing financial proportions for anticipating financial distress have pulled in a lot of consideration of the scholastics and also economic and financial institutions.[6]

Mohamed EL Kourdi et al. have proposed this system in which Naive Bayes (NB) which is a factual machine learning algorithm is utilized to order Arabic web documents. This system utilizes K-Nearest Neighbor for estimating values that will maintain a strategic distance from financial distress and bankruptcy. In the current review k-Nearest Neighbor characterization technique, have been examined for economic estimating. Lately, after the situation of worldwide [7] financial emergency, the quantity of bankrupt organizations has risen. Since organizations' financial distress is the principal phase of bankruptcy, utilizing financial proportions for anticipating financial distress have pulled in a lot of consideration of the scholastics and economic and financial institutions.[3] Kevin Beyer et al. have proposed this system that tries to explain what happens.

III. PROPOSED ARCHITECE

Supervised Learning algorithms learns the pattern fromPre-existing data and try to predict new result

based on the previous learning. ML algorithms are used to identify existing data like probability-based, function-based, rule-based, tree based, instance-based, etc. According to our model patients are required to provide their medical data for successful diagnosis of their diabetes test. The admin of the system will
Then choose one of the two appropriate algorithms available. Thus, after using the system, the prediction will be done whether the patient is diagnosed with diabetes or not. If the patient is found diabetic expert recommendations would be provided to the patient so that he/she can recover from diabetes.

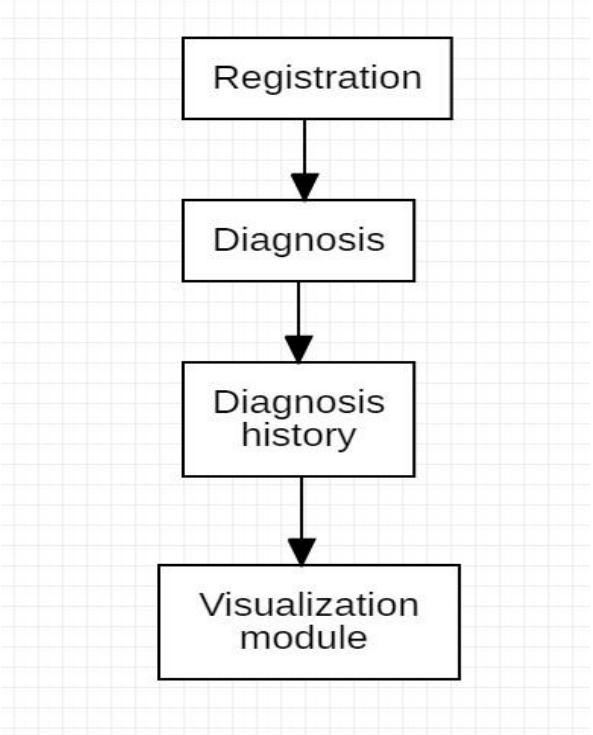


Fig. 1. Flowchart of Diabetes prediction Model

IV. METHODOLOGIES

A. K- Nearest Neighbors (KNN)

K-Nearest Neighbors is a supervised learning algorithm,
K- means a number of a vector. The working methodology of KNN is pretty simple, it's predict based on the value of Kparameter. Graphical representation of K Number Nearest Neighbors are depicted in following Fig. 3.

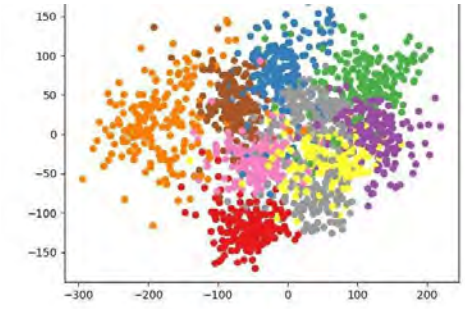
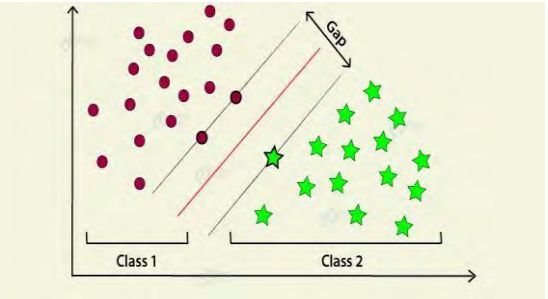


Fig. 3. Graphical representation of KNN working Procedure

B. Support Vector Machine (SVM)

Support vector machine is a supervised learning algorithm mainly used for classification problem. Over fitting derives an ML model to misclassify the data from a given dataset. In that case SVM can prevent over fitting nature from samples data and produce better accuracy [8][9]. SVM possess a linear hyper plane with a margin which divides the dataset



into positive and negative samples [8][9].

Fig. 2. Graphical representation of SVM working procedure

C. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is considered as the method to represent human neuron in a Mathematical way by reflecting its learning and generalization abilities. ANN model can scale up a highly nonlinear system in which the
relationship between the variables is unknown or very complex [11].

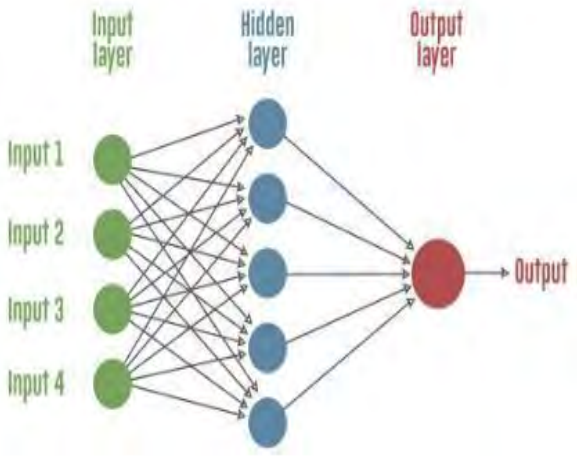


Fig. 4. General Structure of Artificial Neural Network with two hidden layer

A neural network is consist of various neurons followed by some layers. Nodes represent the structure of human neuron as like dendrite and axon where linked between nodes represents as axon with the weighted connection. The total structure of the neural network is shaped by input layer, one or more hidden layers, and the output layer where *ith* neuron represents a connection with *jth* neuron of total structure and
Wij represents as the strength of the link between neurons. The nodes of the structure of an ANN take

inputs (features) then send them to next hidden layer through some sort of weighted
link where *ith* nodes send data to *jth* nodes for processing and calculating the weighting sum and totaling a bias term (θ_j) V. RESULTS

In this section we will discuss regarding our results which we have achieved after experimental design. Following TABLE I. represents an insight description of our Pima Indian Dataset. This dataset is mainly based on the females those were living at Pima Indian heritage. Following 8 features ***(a-h)*** of Pima Indian dataset helps us to predict the diabetes of any Individuals with the help of our proposed methodologies.

a) Numbers of time Pregnant
b) Glucose Test
c) Blood Pressure
d) Triceps skinfold thickness
e) 2-Hour Serum Insulin
f) Body Mass Index
g) Diabetes Pedigree function
h) Age

Class	Attribute Number
Pregnancy Count	1
Glucose concentration in plasma	2
Blood pressure (diastolic, mm Hg)	3
Thickness of triceps skin fold (mm)	4
2-Hour serum insulin (μ U/ml)	5
Body mass index	6
Pedigree function of diabetes	7
Years of age	8

Following TABLE II depicts the average accuracy rate varies with training dataset size. We have experimented with different size of training data for SVM, KNN, GNB and
ANN. Without Min Max Scaling (MMS) it shows an average accuracy of 76.25% with Gaussian Naïve Bayes Algorithm.

```
In [31]: # Evaluating using accuracy_score metric
         from sklearn.metrics import accuracy_score
         accuracy_logreg = accuracy_score(Y_test, Y_pred_logreg)
         accuracy_knn = accuracy_score(Y_test, Y_pred_knn)
         accuracy_svc = accuracy_score(Y_test, Y_pred_svc)
         accuracy_nb = accuracy_score(Y_test, Y_pred_nb)
         accuracy_dectree = accuracy_score(Y_test, Y_pred_dectree)
         accuracy_ranfor = accuracy_score(Y_test, Y_pred_ranfor)

In [32]: # Accuracy on test set
         print("Logistic Regression: " + str(accuracy_logreg * 100))
         print("K Nearest neighbors: " + str(accuracy_knn * 100))
         print("Support Vector Classifier: " + str(accuracy_svc * 100))
         print("Naive Bayes: " + str(accuracy_nb * 100))
         print("Decision tree: " + str(accuracy_dectree * 100))
         print("Random Forest: " + str(accuracy_ranfor * 100))

         Logistic Regression: 71.42857142857143
         K Nearest neighbors: 78.57142857142857
         Support Vector Classifier: 73.37662337662337
         Naive Bayes: 71.42857142857143
         Decision tree: 68.18181818181817
         Random Forest: 75.97402597402598
```

Fig. 5. Implementation of different algorithms on the Pima Indian Diabetes dataset.

Here we have preprocessed diabetes dataset and applied different machine leaning algorithms .Then we found that KNN gives highest accuracy with 78.56% .This implementation is done in Anaconda.

Fig. 6. Confusion Matrix shows the result.

Web Page Implementation

Fig. 7. Analysis Form where user will get prediction

This is the Analysis form where patient will enter data and it will predict the disease according to it. Then after this prediction, diagnosis form will open in which recommendations for patient will be shown.

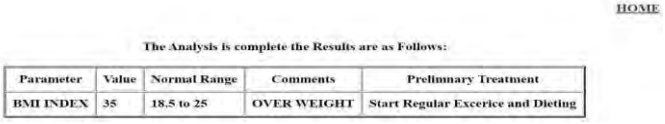


Fig. 8. Recommendation form where diagnosis will be obtained.

V. CONCLUSION

In this paper, we proposed a web based application for the successful prediction of Diabetes Diseases. From different machine learning algorithms Artificial Neural Network
(ANN) provide us highest accuracy with Min Max Scaling Method on Indian Pima Dataset. As we have proposed and developed an approach for diabetes disease prediction using machine learning algorithm, it has significant potential in the field of medical science for the detection of various medical
Data accurately. In near future our focus is to use a deep learning model and prepare a Location based Dataset from medical data for the successful prediction of diabetes disease.

References

[1] Samrat Kumar Dey , Ashraf Hossain "*Implementation of Web Application to Diabetes Disease: An Approach Using Machine Learning Algorithm*" , 2018 21 Conference of Computer and Information Technology (ICCIT), 21-23

[2] Deeraj Shetty, KishorRit, Sohail Shaikh, Nikita Patil, "*Diabetes Disease Using Mining",* 2017 International Conference on Innovations inInformation, Embedded and Communication Systems (ICIIECS)

[3] Pahulpreet Singh Kohli, Shriya Arora, "*Application of Machine Learning in Disease Prediction*", 2018 4th International Conference on Computing Communication and Automation (ICCCA)

[4] https://ieeexplore.ieee.org/document/7154738

[5] https://www.kaggle.com/uciml/pima-indians-diabetes-database

[6]https://developer.mozilla.org/enUS/docs/Learn/HTML/Howto/Use_Java

Improved Prediction Using Modified K-Means

Dr.AnandKhandare, AnjaliThomas, KinjalBhatia, MahimaChoudhary
Thakur College of Engineering And Technology
Mumbai, India
anand.khandare@gmail.com, thomasanjali07@gmail.com, bhatiakinjal98@gmail.com, mahima.choudhary10@gmail.com

Abstract—**Clusteringis oneof themostcommonexploratory dataanalysistechniqueusedtogetan intuitionaboutthe structureofthedata.Wetryto findhomogeneoussubgroups withinthedatasuchthatdatapointsineachcluster areas similaras possible.Clustering the databeforepredictionwould giveusaclearideaofhowto go aboutthedata.Predictionafter implementingthistechniquewouldgive reducederrors.Asa clustercontainsmorehomogenousdata,usingapredictoron thatclusterwillgiveappropriateresult. Predictionbasedon clusteringdependsontheaccuracyofclustering.Soto notgive intodrawbacksoftraditionalk-meansalgorithminclustering thedata,weusea2layeredk-meansalgorithmto improvethe formationofclusters. 2 layeredk-meansdealswiththeoutliers andnoisydataby subclusteringtheclusterandthusreducing errors intheclusters. In this paper, we areapplying the 2 layered k-meansalgorithminimprovingtheaccuracy ofthe clusters which inturn will helpinimproving theprediction accuracy.Thispapertakesyouthroughthestepsof2layeredk-meansalgorithmandclusteringthedata,and benefitsof clusteringbeforeprediction.Our experimentalresults demonstratethatour schemecan improvetheaccuracyofthe prediction.**

Keywords—clustering,sub-clustering,prediction, k-means, accuracy.

I. INTRODUCTION

Clustering isconsideredanunsupervised learning methodsincewedon'thave thegroundtruthtocomparethe output of the clustering algorithm to the true labels to evaluateitsperformance.*K*-meansalgorithmis asimpleand easy clustering methodwhichcanefficientlyclassifyalarge numberofcontinuousnumericaldataofhigh-dimensions. K-meansalgorithm isaniterativealgorithmthattriesto partition the dataset into*K* pre-defined distinct non-overlapping subgroupswhereeachdatapoint belongstoonly onegroup.Ittriestomake theinter-clusterdatapointsas similar as possible while also keepingtheclustersas differentaspossible.Itassignsdatapoints toaclustersuch thatthesumofthesquareddistancebetweenthe datapoints and thecluster'scentroidisattheminimum.The less variationwehavewithinclusters,themorehomogeneousthe datapoints arewithinthesamecluster.

However,itisvulnerabletooutliersandnoisydata,andit spends muchexecutivetimeinclassifyingdatatoo.Noisy data,outliers,andthedatawithquitedifferentvaluesinone clustermay reducetheaccuracyrateofdatamatching obtained by a pattern matching system since the cluster center cannot precisely describe the data in the cluster. Hence,thisstudyprovidesa2layer*K*-means algorithmto solveaboveproblems.In this algorithm,eachclusteris subdividedinto sub-clustersand thencombineswiththe traditionalk-meansalgorithmfordataclustering.Thegoal forundergoingaclusteranalysisisthat

1. Getameaningfulintuitionofthestructureofthe datawe'redealing with.
2. Cluster-then-predictwheredifferentmodelswillbe builtfordifferentsubgroups ifwebelievethere isa widevariationinthebehaviors ofdifferent subgroups.

The pre-processingtaskofusingclusteringtogainmore informationaboutthe dataimprovesthe predictionaccuracy. After clusteringour data, we train thepredictor on each clusterseparately ratherthantrainingthepredictoronthe entiredataset.

II. LITERATURESURVEY

After monitoring our dataset wefound that clustering before prediction would give us more optimized results. Herethe predictionisdependentonclusteringofourdata. Duetothedrawbacksofourtraditionalk-meansalgorithm inclusteringdata,wesoughtto modify howthedatapoints are clustered.Hencethe sub-clustering techniquewas proposed.Aftermodifyingourtraditionalk-meanswe are applyingthatmodifiedk-means algorithmto improvethe prediction of sales of diamonds. Modified k-means algorithmisusedinclusteringthe datainsimilarcategories. Clustering the data before predicting improves the predictionaccuracy.Hencewefocus,moreonthe clustering techniquesas itiscrucialinimprovingprediction.Oncethe clustersareformedwe applyapredictoralgorithmtopredict thesalesofdiamonds

III. SUB-CLUSTERING

K-meansalgorithminitializescentroidsby firstshufflingthe datasetandthenrandomlyselecting kdatapointsforthe centroids. However oneofitsdrawbacksis thatinitialcentroidpointsareselectedrandomlybecauseof whichalgorithmhas tore-iteratenumberof times,andalso accuracy ofthek-meansalgorithm dependsmuchonthechosencentralvalues.Hence,wechoose the initialcentroid points.A mostdiscrepantclustercentermethodisusedto selecttheinitialclustercenter.First,thedatais normalizedThe normalizationisusedtogetbetterresultsintheformof findingdistanceto haveexactcentroidand to removenoise datawhichisnot needed.

For taking two discrepant data points as initial cluster centers, we calculate the Euclidean distance of each data point from the origin. The data point with least distance from the origin is taken as C1 and the data point with largest distance from the origin is taken as C2. After that, it computes the data: C3 which is farthest from C1 and C2, C4 which is farthest from C1, C2 and C3, Ck which is farthest from C1, C2, …, and Ck-1, where C1, C2, …, and Ck are in S and are considered to be the initial cluster centers of the K clusters

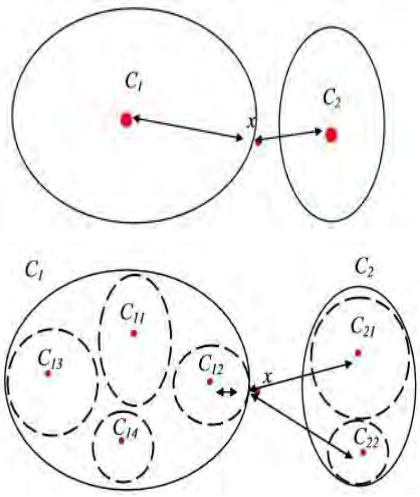


Figure1:Thereare2 clustersC1andC2

Thefigure aboveshowsthatbyusingtraditionalk-means algorithm,datapointxwillbelongtoclusterC2 asthe distancebetween xandcluster centerC2 islessthanthe distancebetweenxand theclustercenterC1.Butthedata pointxshouldbelongtoclusterC1asthe distancebetween theclusterC1 islessthanthedistancebetweenxandthe clusterC2.So toovercomethisdrawbackwesub-clusterthe clusterswhereeachclusterisdividedintosub-clusters.C1andC2 areclusteredintosub-clutered Sub-clustering algorithmusestraditionalk-meansalgorithm todivideeachclustersformedintog sub-clusterswith Cg1,Cg2,....,Cgk.As showninthe figure,aftersub-clustering theclusterswecansay thatxbelongstoclusterC1because thedistancebetweenxandsub-clustercenterC12 islesser than distance calculated between xandanyother cluster center.Thesub-clusteris notonly closesttothedatapoint, butalsobelongstotheclusterCG,andthenthedatapoint is classifiedintoC12.

IV. CLUSTERINGBEFOREPREDICTION

Clusteringisprobablythemost usedexploratorydata analysistechnique acrossdisciplinesandis frequently employedtogetanintuitionaboutthe structureofthedata, for findingmeaningfulgroups,alsoforfeatureextraction andsummarizing.Givenadataset,clustering itgivesa compressedrepresentation.This canbethoughtofasgiving thedatatoan operatorasinput (k-means)thatgives an outputofthesamedatabut taking muchfewerbitsto represent it. This transformation tells us something interestingaboutthedataanditsstructure whichcouldbe exploited toimprovethepredictive power.Onepotential wayofdoingsoisby trainingapredictoroneachcluster separately ratherthantrainapredictorontheentiredataset. This ideaofusingclusteringcoupledwithsimplepredictors beat more complex methods such as Support Vector Machines and Random Forests. To traina "prediction model",thefollowingsteps arefollowed:1.Clusterthe training dataintokpartitions2.Foreachpartitiontraina classifier/predictorusingthepoints insidethatclusterasits training set.3.Eachsuchpredictor representsa modelofthe cluster,andhenceis calledtheclustermodel

V. CONCLUSION

Ifweareusingsub-clusteringasourpreprocessingtask for prediction, then the accuracy of clusters isIf we are using sub-clustering as our pre processing task for prediction, then the accuracy of clusters iscrucial in predicting. The two-layer K-means algorithm isused to strengthen the accuracy of data clustering. In thispaper,sub-clustering algorithm is used to improve traditional k-means algorithm so that accuracy of prediction can be improved. It can give a better accuracy rate of resultsthan the methods used without clustering.

References

[1] Sujata R. Yadav, Balaji G. Hogade, ImdadaRizvi,andSarikaYadav,"Extraction of Vegetation using modified k-meansClustering,"Information and Communication Technology forIntelligent Systems Proceedings of ICTIS 2018, vol. 2, pp. 391-398.

[2] G. Poornalathaand Prakash S. Raghavendra," Web User Session Clustering Using Modified K-Means Algorithm" A. Abraham et al. (Eds.): ACC 2011, Part II, CCIS 191, pp. 243–252, 2011. Springer-Verlag Berlin Heidelberg 2011

[3] JiHoon Kim, Sanghun Lee, GangSeong Lee, YoungSoo Park, and YeongPyo Hong," Using a Method Based on a Modified K-Means Clustering and Mean Shift Segmentation to Reduce File Sizes and Detect Brain Tumors from Magnetic Resonance (MRI) Images", Springer Science+Business Media New York 2016

[4]Shubhendu Trivedi, Zachary A. Pardos, and Neil T. Heffernan," TheUtility of Clustering in Prediction Tasks"

[5] Chen-Chung Liu, Shao-Wei Chu2, Shyr-Shen Yu, and Yung-Kuan Chan," A Modified K-means Algorithm - Two-Layer K-means Algorithm", 2014 Tenth International Conference on IntelligentInformation Hiding and Multimedia Signal Processing

[6]VikasVerma, Shaweta Bhardwaj and Harjit Singh," A Hybrid K-Mean Clustering Algorithm for Prediction Analysis", Indian Journalof Science and Technology, Vol 9(28), July 2016.

[7]Arpit Bansal, Mayur Sharma, and ShaliniGoel," Improved K-meanClustering Algorithm for Prediction", International Journal ofComputer Applications (0975 – 8887) Volume 157 – No 6, January2017

[8]AleksandarLazarevic, RamdevKanapady, Chandrika Kamath, VipinKumar, and Kumar Tamma," Localized Prediction of ContinuousTarget Variables Using Hierarchical Clustering", 2014 TenthInternational Conference on Intelligent information Hiding andMultimedia Signal Processing.

[9]SatishChaurasiya , Dr.Ratish Agrawal," K-Means Clustering WithInitial Centroids Based On Difference Operator", InternationalJournal of Innovative Research in Computer and CommunicationEngineering,Vol 5,Issue 1,January 2017.

Beauty Products Recommendation System Using Content-Based Filtering

Kanika Jain,Tanya Gupta,Deepa Gupta,Dr.Harshali Patil
Computer Department ThakurCollege of Engineeringand Technology, Mumbai
Indiakanikadajain@gmail.com,tanyajaygupta@gmail.com,Indiadeepaguptah@gmail.com
Indiaharshali.patil@thakureducation.org

Abstract **--In the presentglamorousand technologically advancingworldwhereeveryonewishestolookbeautiful – technologycomes handy.Dueto thefastdevelopmentof technology,therehasalsobeen arapid growth in thebeauty industry. Havingsomany products withmultiple specifications makeit difficultforthebuyerstochoosethe righttype ofproductaccordingtotheirneeds.Whilemostwebsitesrelated tobeautyproductsprominentlydisplayratingsand reviewsfor products, theoverallsetofcriteriatosearchfor theseproducts isnotveryrobust.Forinstance,siftingthrough reviewsto find theidealbeautyproductthatmatchestheneedscan often lead to an exorbitant amount of hours of tedious reading through several non-informative reviewsbeforeamatchis found.Because ofthistryingoutanewcosmeticproducthas becomemorethandifficultassometimestheyendupgiving skintrouble.**

Thesolutiontosolvethisproblem istotakeadvantageofthe vastamountoftext-based reviewdata and Content-Based FilteringMethodtobuildabasicrecommendationsystem for makeup, skin-care andhair-care products. The system alsousesSocial MediaMiningtoevaluate thereviews and givean average ratingtoa product.Therecommendation system aims to facilitatethecustomersto taketheright decision in determining which producttobe used in accordancewith the requiredspecifications ofthe users.

Keywords--** **beautyproducts,skin-care,cosmetics,and content- basedfiltering,socialmediamining

I. INTRODUCTION

Intoday'sglobalization eramanyneedsassociatedwith technologyhavearoused.Mostofthe latesttechnologies havecomeintousebecause ofthetimesandtheneeds.In almostallaspectsoflife,information technologytodayhas becomeveryhelpfulin supportinghumanlife.Technological developments suchastheinternethavemadeiteasierfor peopletosearchforinformation [1].Current technologyis developing veryrapidly.Thisrapiddevelopment of technology hasalsoaffectedtherapiddevelopment ofthe beautyindustrytoday.

BasedonthedatafromtheMinistry ofIndustry (2016), theindustrialmarketgrowthaveraged 9.67%peryearinthe lastsixyears(2009-2015). Itisestimatedthatofthetotal marketbeautymarketisapotential marketforentrepreneurs industry.SigmaResearch Indonesia conductedresearchin Indonesiainsegmentsof1200womenagedbetween15to 55years.These studiesrevealedthat therearesevenalfactors totakeintoconsideration regardingwomeninpurchasing beautyproducts.Thebiggestfactoristhepercentageformula thatmatches thefacialskin (79.4%),followed bylong-lastingproducts (67.4%),lightweightformula(62.2%),as wellascolorselectionwithpercentage pricerespectively above50%.

(a)

(b) (c)

Fig.1. Examples (a) Make-up Products (b) Skin-care Products (c) Hair-care Products

Arecommendation systemisamodelapplicationofthe resultsoftheobservationsonthecircumstancesandwishes ofthecustomer[4].Therefore,arecommendation system requiresamodelofpreciserecommendations thatare recommendedinaccordancewiththewishesofcustomers, aswellasenabling customerstotaketherightdecision in determiningwhichproductswillbebestfortheiruse.For

this,content-based filteringisdonebasedoncommon approaches betweenitemprofilesanduserprofiles[5].The systemisexpectedtohelpsomeone whoisgoingtobuy cosmeticstoconformtotherequiredspecifications.

Similarly,wewillbefollowing thisabovesystem and improvingit byexpandingtheavailabledatasetbyscrapping variousbeautyproductandreviewwebsites. Taking advantage ofthevastamountoftext-basedreview dataand Content-BasedFilteringmethodwe aimto builda basic recommendationsystemformakeup,skin-careandhair-care products. ThesystemalsousesSocialMediaMiningto evaluatethereviewsandgiveanaverageratingtoaproduct. Therecommendationsystem aimstofacilitatethecustomers totaketherightdecisionindeterminingwhich producttobe usedinaccordancewiththerequiredspecifications ofthe users.

II. LITERATURE SURVEY

Asthetechnologyadvances,moreinformation iseasily availabletosolveaproblem.Applicationsutilizing a recommendationsystemarealreadygrowingrapidlyasitcan beseeninvariousfields suchaseducation, business, automotive, agriculture, marketingandsales. Recommendation systemsaretypicallyusedtomakea decision.AccordingtoSon,MahmudandSetiawanintheir paper RecommendationSystem forElectivesof Students withContent-Based FilteringandCollaborativeFiltering (2015)arecommendation systemwillprovidedifferent recommendations toeachuser,notjustprovidealistofthe itemsmostindemand,butratherprovideadviceonitems thatmaybeappropriate forusers[3]. Thatis,eachuserwill getdifferentrecommendations, accordingtotheprofileand userinterest.

Maharani andGunawanintheirpaperCar Recommendation SystemBasedonDemographicand Content-BasedFiltering(2015)madearesearchaboutthe carselectionbasedonDemographic andContent-Based Filtering [1].ThisresearchcombinesmethodsDemographic andContent-Based Filteringinordertomake recommendationsthatreallyfit theprofileandpreferencesof thebuyers.DemographicFilteringmethodis commonlyused tohandle thenewlyregistered userorusers whodonothave theprofileorrecordsof transactionsoractivityinthesystem. Inthisstudy,recommendations areobtainedbasedonthe similaritybetweenuserpreferencesandspecificationsofcars availableinthe system.

Research byWayandAnggriawan inthepaper Recommendation System Package Se Malang Hybrid Method UsingContent-Based and Collaborative(2013) madearecommendation systemforthemarketingof merchandisesales[4].Thepurposeofthisstudyisto provide thebestinformation forsalesinmarketingtheirproductsto shopsorstores. Therearetwoapproaches tofiltering information thatiscollaborativefilteringandcontent-based filtering.Collaborative Filteringisbasedonthesimilarity between userswhooptintothesystemwhichisfurther divided intotwo, User-Based CollaborativeFiltering andItem-BasedCollaborativeFiltering.ContentBasedFiltering ontheotherhandrecommendsan itemtoa userbasedonthe descriptionoftheitemandtheprofileofthe user'sinterest.

GuoshuaiZhao,XuemingQian,XiaojiangLei&Tao MeiinthepaperServiceQualityEvaluation byExploring SocialUsers'ContextualInformation (2016)proposeda modeltoconductservicequalityevaluation byimproving overallratingofservicesusinganempiricalmethodology [6]. Theyused theconceptof user rating'sconfidence,which denotesthetrustworthiness ofuserratings.First,entropyis utilizedtocalculateuserratings'confidence. Second,they furtherexplorespatial-temporal featuresandreview sentimentalfeatures ofuserratingstoconstraintheir confidences. Last,theyfusethemintoaunifiedmodelto calculateanoverallconfidence, whichisutilizedtoperform servicequality evaluation.Extensiveexperiments implemented onYelpandDoubanMoviedatasets demonstratetheeffectivenessofthismodel.

Asubsequent research byPattyJoanna(2017)madea recommendation topurchasethesystemdesigncosmetics usingContent BasedFiltering [2]. Thepurpose ofthisstudy istoestablishasystemofcosmeticrecommendations for someonewhowillbuycosmetics thatsuityourneeds according tohisskincolor.Inthisstudy,ContentBased Filteringisdone basedoncommonapproaches,itemprofiles anduserprofiles.Then measuresofthesimilarityoftheitem profilesanduserprofiles arecalculatedusingCosine Similarity.

III. WORKING

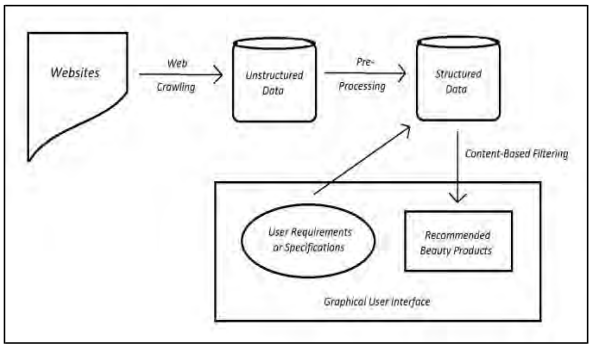


Fig.2. BlockDiagram

Asunderstood bythediagram WebCrawlingisusedto extract datafromvariouswebsites.Thisdataisunstructured andthereforePre-processing iscarriedouttoobtain structured data. Following therequirements andspecifications givenbytheuserwhileperformingContent-BasedFilteringonstructured data,beauty products are recommended accordingly. TheblocknamedasGraphical UserInterfacedepictsthecomponentsshowntotheuserand bywhichthe userinteractswiththe systemaswell.

Firstwestartbydoing WebCrawling usingPythonon various beautyproductwebsitestoextractinformationabout productslikebrand,price,finish,skintype,etc.andcalculate ratingusing reviewsacquiredfrom thosewebsites.Using WebCrawlingwecanaccessandextractlargeamounts of informationfromawebsite,whichcansaveahugeamount oftimeandeffort.Theycrawlonepageatatimethrough a websiteuntilallpageshavebeenindexed.Webcrawlershelp

in collecting information about a website and the links relatedtothem,andalsohelpinvalidatingtheHTML code andhyperlinks.Tobuildarecommendationsystembasedon reviewcontentsfromdifferent products,weextracted information ofproductsandreviewsbycrawling2levels down.

Fig.3.WebCrawling

ThenwedidPre-Processing forconvertingthe unstructured datatostructureddata.Datapreprocessingisa dataminingtechniquewhichisusedtotransformtheraw datainausefulandefficient format.StepsinvolvedinPre-Processingare:-

1) *DataCleaning*
2) *DataTransformation*
3) *DataReduction*

LastlyweperformedContent-BasedFiltering to recommendbeauty products according tothefiltersselected bytheusers.Filters such asprice,brandname,rating,finish,

ACKNOWLEDGMENT

Wewouldliketoacknowledgethecontribution ofour guideDr.Harshali Patilforguidingusthroughthistechnical paper.Wealsowhole-heartedlyacknowledgetheinstitute andourdepartmentforgivingusthisopportunit

References

[1] Maharani andGunawan, "CarRecommendation SystemBased on DemographicandContent-BasedFiltering",TelematicsJournal,Vol 2,No.9, pp. 64-68, 2015.

[2] Patty Joanna, "Electoral System Design Recommendations for Purchase of Cosmeticsusing Content-BasedFiltering", pp.1-7, 2017.

[3] Son, Mahmudy and Setiawan, "Recommendation System for ElectivesofStudentswithContent-BasedFiltering andCollaborative Filtering",RepositoryPTIIKUB'sStudentJournal,Vol1,pp.1-11, 2015.

[4] Wahyo and Anggriawan, "Recommendation System Package Se Malang Hybrid Method using Content-Based and Collaborative- BasedFiltering",ScientificJournal ofInformationTechnology and ASIA,Vol1,No.9, pp.6-13, 2015.

[5] SharmaandGera,"A SurveyofRecommendationSystem: Research Challenges",InternationalJournal ofEngineeringTrendsand Technology(IJETT),Vol5,No.4, pp. 1989-1992,2013.

[6] GuoshuaiZhao,Xueming Qian,Xiaojiang Lei,andTao Mei,"Service Quality Evaluation by Exploring Social Users' Contextual Information,IEEE,2016..

[7] JoannaCristyPatty,ElikaTheaKiranaandMadeSandraDiamond Khrismayanti Giri,"RecommendationsSystemforPurchase of

Recommendationsaredisplayedbasedonthespecifications enteredbythe user.

IV.CONCLUSION

Because ofenhancedtechnology,therehasbeenagreat amountofimprovement,but we arenot ableto use the enhancedtechnologyefficiently. Beitanykindofindustry - automobile,beauty orelectronics-therearemanyproducts andmany variationsthatweareactuallyunawareofandend upcompromisingourneedsbybuyingthe knownproducts.

Through oursystem,wehavetargeted thebeauty industryasitisoneoftheindustriesthathasagreatamount ofrisk.Thisisbecauseanypersonbeforetryinga newproducthasathreat ofitssideeffects. Soweaimtomakea systemwheretheusercanfindaproductaccording totheir requirementsand specificationswithoutcompromisingon theirneeds.

andskintypeareprovided byustoallowright recommendationofthebeautyproductstothe users.

Fig.4.SkinTypes

RecommendedSystemSelectionPurchasingcosmeticsi s asystembuilttoprovidecosmeticrecommendation in accordance withtherequirementsdesiredbytheuser.

Cosmetics using Content-BasedFiltering,InternationalJournalof ComputerEngineeringandInformationTechnology,Vol 10,No.1, pp.1-5, 2018.

[8] R. Zafarani, "Social Media Mining: An Introduction", Cambridge UniversityPress, 2010.

[9] JiwonJeong,"ForYourSkin Beauty: Mapping CosmeticItemswith Bokeh",URL - https://towardsdatascience.com/for-your-skin-beauty- mapping-cosmetic-items-with-bokeh-af7523ca68e5, 2018.

Emotion Detection Using Facial Expression

Mrs. Vidyadhari Singh, Abhimanyu Singh, Shashikant Maurya, Dhanesh Akolu
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
abhisingh03101997@gmail.com, shashimourya1@gmail.com, dhaneshakolu9@gmail.com

Abstract **:-Implementing system for emotion detection from from facial expression With combination of neural network for the recognition of different facial emotions (happy, sad, angry, fear, surprised, neutral etc..). Humans are capable of producing thousands of facial actions during communication that vary in complexity, intensity, and meaning. Purposed system depends upon human face as we know face also reflects the human brain activities or emotions. In this project neural network has been used for better results.**

I. INTRODUCTION

This algorithm automatically extracts features and their motion information, discriminate subtly different facial expressions, and estimate expression intensity is showing how Emotion recognition using brain activity performs its task. We know that emotions play a major role in a Human life. At different kind of moments or time Human face reflects that how he/she feels or in which mood he/she is. Humans are capable of producing thousands of facial actions during communication that vary in complexity, intensity, and meaning. Emotion or intention is often communicated by subtle changes in one or several discrete features. The addition or absence of one or more facial actions may alter its interpretation. In addition, some facial expressions may have a similar gross morphology but indicate varied meaning for different expression intensities. In order to capture the subtlety of facial expression in non-verbal communication, I will use an exist tings emulator which will be able to capture human emotions by reading or comparing facial expressions.

II. RELATED WORK

Sr.No	Name Of The Author	Method	Techniques
1	Manasa B (2016) et.al	FaceFeature Extraction Method	Region-Based Segmentation
2	Deepika Ishwar (2015) et.al	FaceFeature Extraction Metho d (Using Eye And Mouth) Using BezierCurv e	Skin-color, Region- Based Segmentation
3	A.D.Chitra (2015) et.al	Canny Ed ge Detection	Edge – based segmentation

		Method	
	Xiaoming	ColorSpace,	Edge – based segmentation
4			
		Ed ge	
	CH	Detection	
	EN (2015)	Method	
	et.al		

III. TECHNOLOGY USED

TensorFlow TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications. TFlearn is a modular and transparent deep learning library built on top of Tensorflow. It was designed to provide a higher-level API to TensorFlow in order to facilitate and speed-up experimentations, while remaining fully transparent and compatible with it. TFLearn features include: Easy-to-use and understand high-level API for implementing deep neural networks, with tutorial and examples. Fast prototyping through highly modular built-in neural network layers, regularizers, optimizers, metrics...

• Full transparency over Tensorflow. All functions are built over tensors and can be used independently of TFLearn.

• Powerful helper functions to train any TensorFlow graph, with support of multiple inputs, outputs and optimizers.

• Easy and beautiful graph visualization, with details about weights, gradients, activations and more...

• Effortless device placement for using multiple CPU/GPU.

The high-level API currently supports most of recent deep learning models, such as Convolutions, LSTM, BiRNN, BatchNorm, PReLU, Residual networks, Generative networks... In the future, TFLearn is also intended to stay up-to-date with latest deep learning techniques.

IV. IMPLEMENTATION DETAIL

Steps :

• First, we use haar cascade to detect faces in each frame of the webcam feed.

• The region of image containing the face is resized to 48x48 and is passed as input to the ConvNet.

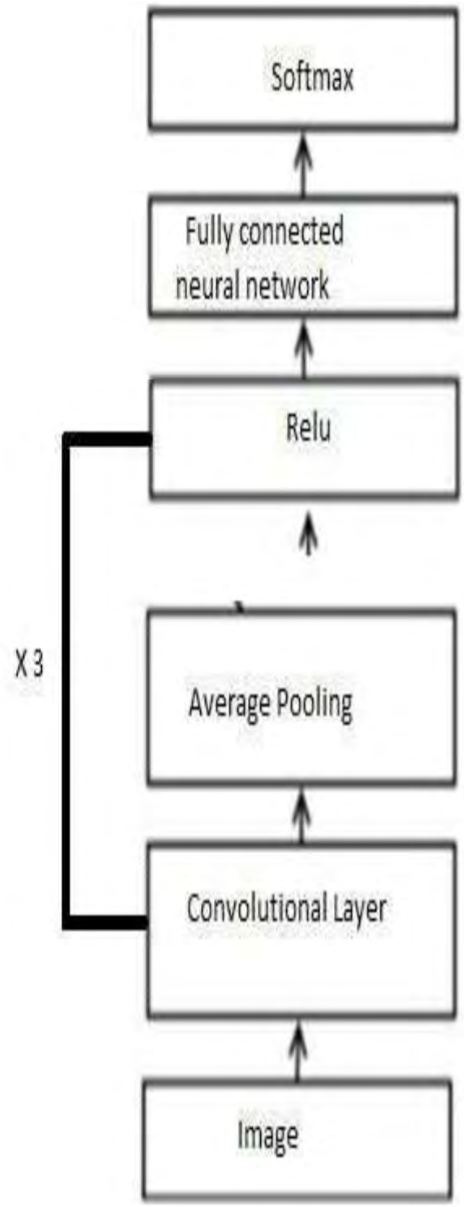
• The network outputs a list of softmax scores for the seven classes. The emotion with maximum score is displayed on the screen.

• A Haar Cascade is basically a classifier which is used to detect the object for which it has been trained for, from the source.

• The Haar Cascade is trained by superimposing the positive image over a set of negative images. The training is generally done on a server and on various stages. Better results are obtained by using high quality images and increasing the amount of stages for which the classifier is trained.

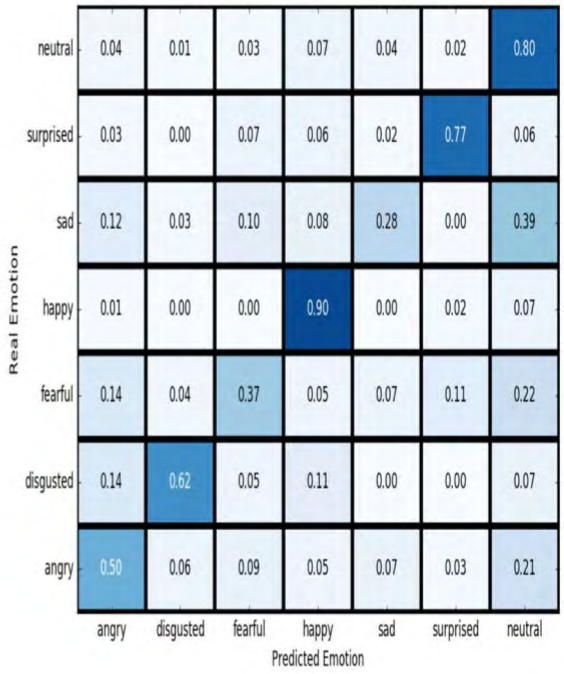
VI. BLOCK DIAGRAM/FLOWCHART

5.1Block Diagram

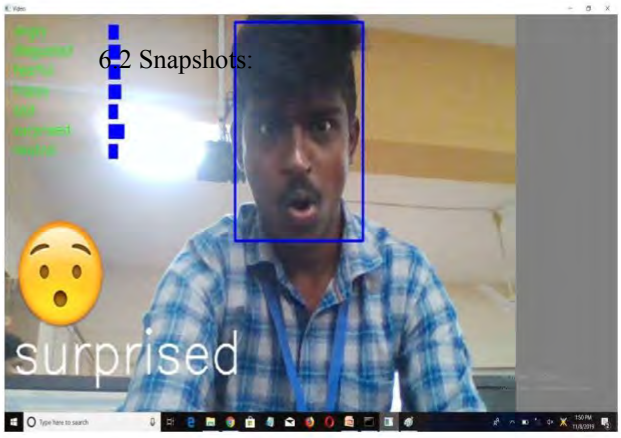


V. RESULT

6.1 Expected Outcome:



6.2 Snapshots:



VII. CONCLUSION:

Sign language gesture recognition as the study found out will be helpful to deaf people in assisting them to interact with computers. All the participants also appeared to be interested, involved and motivated by the system. However, it was not easy for most of the participants to place their hand on the rectangular box, which was considered to be the region of recognition. Hence it is appropriate to recommend that a bigger area is required so that the participants will not have to struggle to place the hand in the region before they start performing the gestures. For the system to be effectively applied in a controlled environment like a deaf classroom then a white background can be placed behind every seat of the computer user to reduce the challenge of complex background noise and improve the background subtraction.

References

1. Miyakoshi, Yoshihiro, and Shohei Kato. “Facial Emotion Detection Considering Partial Occlusion Of Face Using Baysian Network” Computers and Informatics (2011): 96–101.

2. Hari Krishna Vydana, P. Phani Kumar, K. Sri Rama Krishna and Anil Kumar Vuppala. ”Improved emotion recognition using GMM-UBMs”. 2015 International Conference on Signal Processing and Communication Engineering Systems

3. B. Schuller, G. Rigoll M. Lang. ”Hidden Markov model-based speech emotion recognition”. Proceedings. 2003 International Conference on Multimedia and Expo, 2003.

4. Poria, Soujanya; Cambria, Erik; Bajpai, Rajiv; Hussain, Amir (September 2017). “A review of affective computing: From unimodal analysis to multimodal fusion”.

5. Caridakis, George; Castellano, Ginevra; Kessous, Loic; Raouzaiou, Amaryllis; Malatesta, Lori; Asteriadis, Stelios; Karpouzis, Kostas (19 September 2007). Multimodal emotion recognition from expressive faces, body gestures and speech.

7. Machine Recognition of Music Emotion: A Review YI-HSUAN YANG and HOMER H. CHEN, National Taiwan University Balahur, Alexandra; Hermida, JesúS M.; Montoyo, AndréS (1 November 2012). ”Detecting implicit expressions of emotion in text: A comparative analysis”. Decision Support Systems

8. Medhat, Walaa; Hassan, Ahmed; Korashy, Hoda (December 2014). “Sentiment analysis algorithms and applications: A survey”. Ain Shams Engineering Journal.

9. Madhoushi, Zohreh; Hamdan, Abdul Razak; Zainudin, Suhaila (2015). “Sentiment analysis techniques in recent works”. 2015 Science and Information Conference (SAI).

10. Hemmatian, Fatemeh; Sohrabi, Mohammad Karim (18 December 2017). “A survey on classification techniques for opinion mining and sentiment analysis”. Artificial Intelligence

11. Sharef, Nurfadhlina Mohd; Zin, Harnani Mat; Nadali, Samaneh (1 March

2016). “Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data”. Journal of Computer Science.

12. Facial Expression Recognition Using Attentional Convolutional Network. Authors: Shervin Minaee, Amirali Abdolrashidi, Expedia Group,University of California, Riverside.

13. EMOTION DETECTION USING MACHINE LEARNING Authors: Ajay.B.S, Anirudh.C.R, Karthik Joshi.S, Keshava.B.N, Mrs. Asha.N.

Implementing Intelligent Malware Detection Using Deep Learning

Dr Megharani Patil, Manisha Sharma, Ishita Sharma, Siddharth Palepu
Department of Computer Engineering
Thakur College of Engineering and Technology
Kandivali East, Mumbai
tcet@thakureducation.org

Abstract.: **We propose a simple and successful method for visualizing and classifying malware using image processing techniques. Malware binaries are visualized as gray-scale images. Mal- ware coders change small parts of the original source code to produce a new variety. Images can capture small changes and the global structure. Hence, malware variants belonging to the same family appear very similar as images. These images are also different from images of other malware families. In the internet age malware poses a serious and evolving threat to the security making the detection of malware of utmost concern. Security vendors are doing their best to defend against malware attacks but unfortunately with millions of malware discovered monthly they cannot achieve that, thus novel approaches such as deep learning are needed.**

Keywords: deep learning, malware binaries, malware detection, image processing

I. INTRODUCTION

Malware is malicious software to infiltrate the secrecy, integrity, and functionality of a system, such as viruses, worms, trojans, backdoors, spyware. With computers and the Internet being essential in everyday life, malware is a serious threat to their security. It is estimated that one in four computers operating in the India are infected with malware. The threat posed by malware is not only emotional, but financial as well. According to a recent report from cyber crime department, up to one billion dollars was stolen in roughly two years from financial institutions worldwide, due to malware attacks. As a result, the detection of malware is of major concern to both the anti- malware industry and researchers

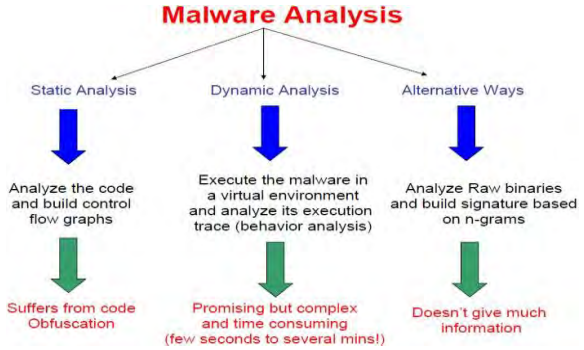


Fig1:Various Techniques of Malware Analysis

In order to pro- tect legitimate users from the attacks, the majority of anti-malware software products (e.g., Comodo, Symantec, Kaspersky) use the signature-based method of detection. Signature is a short string of bytes, which is different for each known malware so that its future examples can be correctly classified with a small error rate. However, this method can be easily evaded by malware attackers through the techniques such as en- cryption, polymorphism and obfuscation. Furthermore, malicious files are being spread at a rate of thousands per day, making it difficult for this signature-based method to be effective. In order to combat the malware attacks, intelligent malware detection techniques need to be investigated. As a result, many researches have been conducted on intelligent malware detection by applying data mining and machine learning techniques in recent years Based on different feature representations, differ- ent kinds of classification methods, such as Artificial Neural Network (ANNs), sup- port Vector Machine (SVM), Naive Bayes (NB), and Decision Tree (DT), are used for model construction to detect malware[1] . Most of these methods are based on shallow learning architectures. Though they had isolated success in malware detec- tion, shallow learning architectures are still somewhat lacking behind with respect to malware detection problems.

Deep learning (DL), a new frontier in data mining and machine learning, is starting to be leveraged in industrial and academic research for different applications (e.g., Com- puter Vision) . A multilayer deep learning architecture is of superior ability in feature learning[1]. More importantly, deep learning architectures overcome the learning dif- ficulty through layer wise pretraining, i.e. pretraining multiple layers of feature detec- tors from the lowest level to the highest level to construct the final classification model. This inspires us to devise a malware detection architecture based on deep learning

II. RELATED WORK & LITERATURE REVIEW

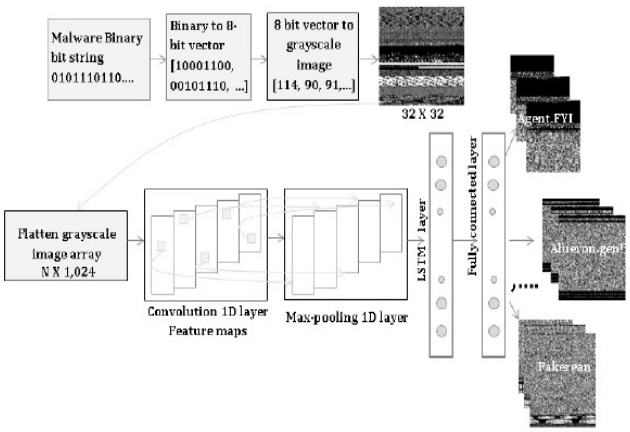
Signature-based method is widely used in anti-malware industry. However, malware authors can easily evade this signature-based method through techniques such as en- cryption, polymorphism and obfuscation. Driven by the economic benefits, the

quan tity, diversity and sophistication of malware has significantly increased in recent years. In order to effectively and automatically detect malware from the real and large daily sample collection, new, intelligent malware detection systems have been devel- oped by applying data mining and machine learning techniques. These intelligent malware detection systems are varied in their uses of feature representations and clas- sification methods. Naive Bayes on the extracted strings and byte sequences was ap- plied in which claimed that Naive Bayes classifier performed better than traditional signature-based method[1].Due to its superior ability in feature learning through multilayer deep learning archi - tecture, deep learning is feasible to learn higher level concepts based on the local fea - ture representations. As a result, researchers have paid much attention to deep learn- ing methods in the domains of natural language processing, computer vision, etc. In recent years, limited research efforts have been made to the malware detection using deep learning. we attempt to explore a deep learning architecture with the CNN model to learn generic features of malware and thus to detect newly unknown malware.

III. PROPOSED METHODOLOGY

Malware is a threat for every modern organization. Attackers and cybercriminals are always coming up with new malicious software to attack their targets. Security ven - dors are doing their best to defend against malware attacks but, unfortunately, with millions of malware discovered monthly, they cannot achieve the target. Thus, novel approaches such as deep learning are needed. Convolutional Neural Networks (CNNs) is a deep learning approach to tackle the image classification problem, or what we call computer vision problems, because classic computer programs face many challenges and difficulties to identify objects for many reasons, including light - ing, viewpoint, deformation, and segmentation.

Fig 2: System Architecture [1]



This technique is inspired by how the eye works, especially the visual cortex function algorithm in animals. CNN are arranged in three-dimensional structures with width, height, and depth as characteristics. In the case of images, the height is the image height, the width is the image width, and the depth is RGB channels. To build a CNN, we need three main types of layers: Convolutional layer: A convolutional operation refers to extracting features from the input image and multiplying the values in the fil- ter with the original pixel values Pooling layer: The pooling operation reduces the di- mensionality of each feature map. Fully-connected layer: The fully-connected layer is a classic multi-layer perceptrons with a softmax activation function in the output layer.

IV. TENSOR FLOW

Libraries**:** It is an open source artificial intelligence library, using data flow graphs to build models. It allows developers to create large-scale neural net- works with many layers. TensorFlow is mainly used for: Classification, Perception, Understanding, Discovering, Prediction and Creation.

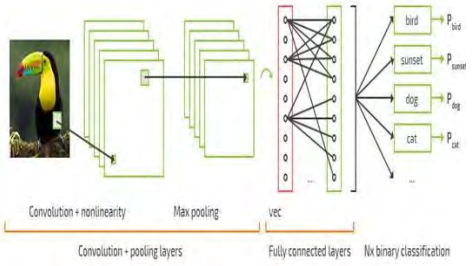


Fig 3: Working of CNN's

4.2 Keras Libraries: Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Tool- kit, Theano, or Plaid-ML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

4.3 Python Script: A plain text file containing Python code that is intended to be di- rectly executed by the user is usually called script, which is an informal term that means top-level program file.

4.4 Fast ai library: Jeremy Howard and Rachel Thomas, founders of fast.ai. fast.ai can be described as a research lab bundled with courses, an easy- to-use Python library with a huge community. Their library wraps popular deep learning and machine learning libraries for common workflows and provides a user-friendly interface

V. CONVOLUTIONAL NEURAL NETWORKS

In this work, CNN network composed of convolution 1D layer, pooling 1D layer and fully connected layer. A CNN network can have more than one convolution

1D layer, pooling 1D layer and fully connected layer. In convolutional 1D layer, the filters slide over the 1D sequence data and extracts optimal features. The fea- tures that are extracted from each filter are grouped into a new feature set called as feature map. The number of filters and the length are chosen by following a hyper parameter tuning method. This in turn uses non-linear activation function, Relu on each element. The dimensions of the optimal features are reduced using pooling 1D layer using either max pooling, min pooling or average pooling. Since the maximum output within a selected region is selected in max pooling, we adopt max pooling in this work. Finally, the CNN network contains fully con- nected layer for classification. In fully connected layer, each neuron contains a connection to every other neuron. Instead of passing the pooling 1D layer fea - tures into fully connected layer, it can also be given to recurrent layer, LSTM to capture the se- quence related information. Finally, the LSTM features are passed into fully con- nected layer. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area. A ConvNet is able to successfully capture the Spatial and Tem- poral dependencies in an image through the application of relevant filters. The archi- tecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.

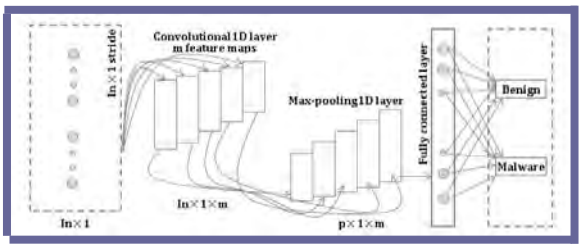


Fig 4:Workflow for malware Detection

1] To achieve the goal of malware detection, in this project we use Convolutional neural networks. (CNN). CNN's utilize the concept of deep learning to extract features from an image.

2] The project starts with the collection of malware and benign binaries and then converting these binaries to 8-bit vector representation (1000 0011…) and thereafter converting these values to image pixel values of 0 and 255. Here 0 represents black while 255 represents white.The values between 0 and 255 represent shades of grey.

3] Thus through this process the malware binaries are converted into a grayscale im- age. The feature extraction of grayscale image helps us to determine whether the given sample is malware or benign.

4] The Pre-processing required in a convolutional neural network is much lower as compared to other classification algorithms.

5] A Conv-Net is able to successfully capture the Spatial and Temporal dependen- cies in an image through the texture pattern, frequencies in image, intensity, or color features, using different techniques such as Euclidean distance, or mean and standard deviation, to generate later feature vectors. In our case, algorithms such as a color lay- out descriptor, homogeneous texture descriptor, or global image descriptors (GIST) can be used. GIST can be accurately calculated using a python library is known as the Pyleargist.

6] The malware is transformed into grey scale images by reading 8-bit unsigned in- tegers. The width of image is defined by file size and height is allowed to vary depend- ing on the width and file size. The malware images are resized to two-dimensional (2D) matrix of 32 × 32 and mapped into 1x1024 size array. Each feature array is nor- malized using L2 normalization.

7] This transformation becomes necessary as the grey scale image that is being used is 2-Dimensional and the convolutional neural network layer is one dimensional, so we need to reduce the dimensions of the image from 2-D to 1-D.Reducing the dimension of the image makes it compatible for passing it to the convolutional layer and thus suit- able to extract features.

VI. RESULTS AND DISCUSSION VII.

This project explores a novel technique of detecting malware through images. It imple- ments detection through images by implementing convolutional neural networks. con- volutional neural networks are a deep learning concept which effectively can capture the texture features of an image to understand whether the given sample is malware or not. convolutional neural networks have been proved to be the most effective tech- nique giving an accuracy of about 98%, which is much higher compared to other ma- chine learning algorithms such as SVM, KNN, Decision trees, which have been earlier experimented with it. This Deep learning approach proves to be a much better than sig- nature-based approaches which were initially used for malware detection. This tech- nique could also prove much useful in case of polymorphic viruses which change

their signatures at every stage of propagation evade detection by signature-based antiviruses.

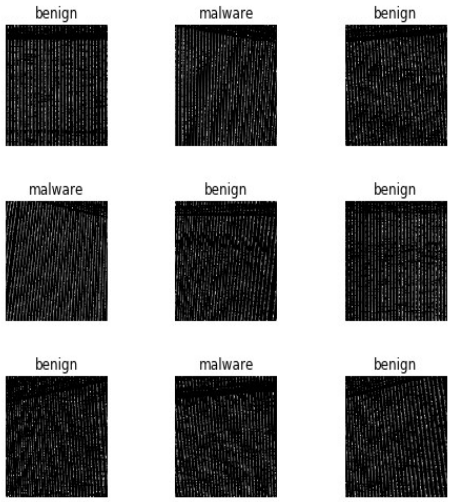


Fig 5: Input dataset consisting of grayscale, malware and benign images

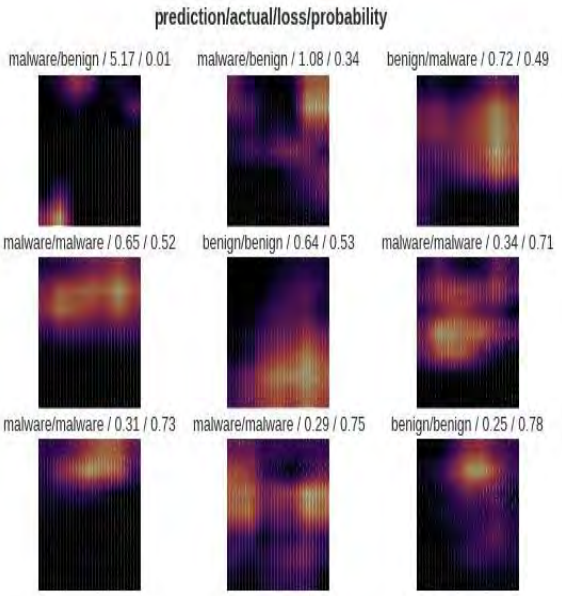


Fig 7: Multiclass classification

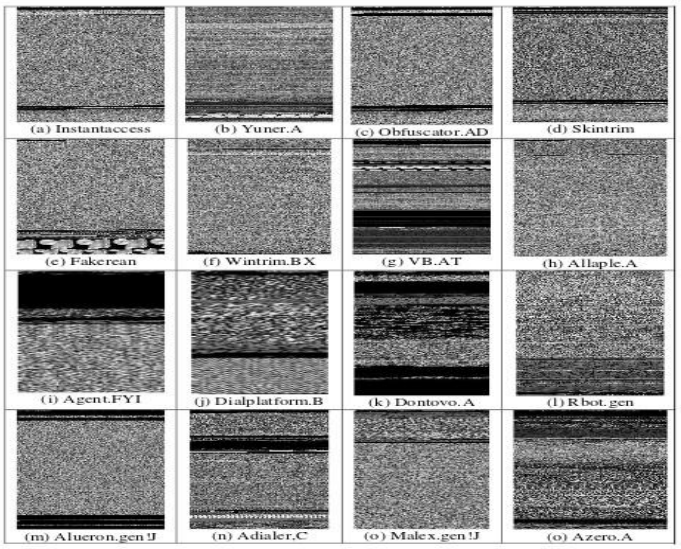


Fig 6: Output Predicted by CNN

6.1 Expected output is to implement multi-class classification of malware families. Image of a variants of a malware belonging to the same family will have similar prop- erties or texture. We aim to capture the texture of the images to classify a incoming im- age into a particular malware family. We use here about 25 malware families in which the incoming image is classified. We implement this system by implementing CNN's

VIII. CONCLUSION

The performances obtained by deep learning architectures outperformed classical MLAs in static, dynamic and image processing-based malware detection and categorization. However, in the dynamic analysis-based malware detection study, the deep learning architectures are applied on the domain knowledge extracted features. This can be avoided by collecting memory dumps for binary files at run time and then memory dump file can be mapped into grayscale image. The deep learning architectures are vulnerable in an adversarial environment. The method generative adversarial network can be used to generate samples during testing or deployment stage. In the proposed work, the robustness of the deep learning architectures is not discussed. This is one of the significant directions towards future work since the mal- ware detection is an important application in safety-critical environment. A single misclassification can cause several damages to the organization.

IX. ACKNOWLEDGEMENT

We sincerely thank our guide Dr. Megharani Patil for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

References

[1] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran and S. Venkatraman, "Robust Intelligent Malware Detection Using Deep Learning," In *IEEE Access, vol. 7, pp. 46717-46738, 2019.*

[2] William Hardy, Lingwei Chen, Shifu Hou, Yanfang Ye∗ , and Xin Li, "A Deep Learning Framework for Intelligent Malware Detection",In *2016 International Con- ference of Data Mining ,ISBN: 1-60132-431-6, CSREA Press*

[3] Wei Wang, Ming Zhu, Xuewen Zeng, Xiaozhou Ye and Yiqiang Sheng, "Malware traffic classification using convolutional neural network for representation learning," *In 2017 International Conference on Information Networking (ICOIN), Da Nang, 2017, pp. 712-717.*

[4] Kolter, J. Z. and Maloof, M. A. 2004. "Learning to detect malicious executables in the wild", In *International Conference on Knowledge Discovery and Data Mining, 470-478*

[5] Torralba, A. Murphy, K.P. Freeman, W.T. and Rubin, M.A. 2003, "Context-based vision systems for place and object recognition", *In International conference on Computer Vision (ICCV).*

[6] B. Klaus and P. Horn , Robot vision , Cambridge, MA, MIT Press, 1986.

[7] Goodall,J.H Randwan H. and Halseth,L 2010. "visual analysis of code security" In *International Workshop on Visualization for Cyber Security(Viz Sec)*

[8] Oliva, A. and Torralba A,2001. "Modeling the shape of a Scene: a holistic repre- sentation of the spatial envelope *In International Journal of Computer Vision,Vol.42(3),145-175*

Implementing System For Object Recognition Using Image Segmentation

Guided By. Ms. ASHWINI PATIL, Group Members Names: RAJ PURAIYA, BALWANT DEO

Abstract: **Object detection with help of image Processing are important and challenging tasks in many computer applications such as vehicle navigation, surveillance and autonomous robot Navigation. Video observation in a dynamic environment, especially for humans and vehicles, is one of the current challenging research topics in computer fields. Dynamic environment is a key technology to fight against crime, public safety, terrorism and for efficient management of traffic. Object Detection in image processing is the first relevant step of information and circumstantial subtraction is a very popular approach.**

Detecting and tracking of human body parts is important in considerate human activities. Intelligent and automatic safety observation systems have become an active research area in

Recent time due to an increasing demand for such Systems in public areas such as airports, underground stations and mass events. It is very difficult for a human machinist to effectively detect events as they happen. Recently computer vision research has to address ways to routinely some of this data, to assist human operators.

Keywords: Object Detection, Convolution Neural Network, Deep Learning, Image Recognition.

I .INTRODUCTION

The Object recognition is the area of artificial intelligence (AI) and other AI implementations to recognize various effects and articles. Object Recognition allows robots and AI programs to pick out and recognize objects from inputs like video and camera images. Google and Microsoft are among the companies working in the area -- Google's Driverless car and Microsoft's Kinect system both use object recognition. The object recognition technology is closely related to social life, object recognition technology is an important branch of computer vision, object recognition technology is along with the modern computer technology, image processing, artificial intelligence, and pattern recognition theory developed a new kind of object recognition technology. To realize the recognition of images, the first to get corresponding image by image acquisition device, so that the digital image.

Object recognition involves a lot of information operation, requiring high processing speed and recognition precision, real-time and fault-tolerance of the neural network in accordance with the requirements of object recognition. Object recognition method, aiming at the limitations of traditional methods, and the complex situations such as images show different state, in the process of image processing algorithm for the image segmentation study and its improvement. Progressions curtail or maybe break. Numerous investigators presently trust that these problems range unit partly related with diet consumption. The project aim is to develop associate degree application for estimating nutrition calories and improve people's consumption conducts for health-care. Provided that employers, patients with suitable and smart explanations that enable them live their food consumption and gather nutritional info area unit the primary valued visions near long run interfering and no-hit treatment programs and it are often conjointly employed in agricultural sector, grocery store.

II. RELATED WORK

The robotic fruit harvesting system is developed with the help of fruit detection algorithm using multiple structures identical intensity, color, alignment and edge of the fruit images. With the help of improved multiple feature-based algorithm, the detecting effectiveness is attained up to 90% for various fruit items [1]. For the exploration of the image FFB, the expansion of out-of-doors image inspection of oil palm fruit fresh bunches (FFB) are essential. The software examination generates the accurate prototypical and connection component amongst the light intensity in kin to value of FFB from RGB element of image occupied [2]. The on-line valuation of the superiority of fruits the calculation of the effectiveness of these methods concerning the next superiority facets hereby size, color, stem position and recognition of outer flaws is offered [3]. The main stages of the pipeline are segmentation of items from background, feature extraction mainly based on color, and classification with Gaussian Bayes classifier [4]. An automatic spherical fruits recognition system in the natural conditions facing difficult situations such as shadows, bright areas, occlusions and overlapping fruit [5] Convolutional neural Network achieved ample improved than did outdated approaches By means of handcrafted features. Complete comment of competent convolution kernels, we inveterate that color structures are vital to nutrition image identification [6]. Defined nutrition identification consuming a minor dataset, which was proposed to be second handed in a Smartphone based food classification scheme [7]. Which identifies unhealthy foods beginning cartridges of eating and guess meal calories created on identifying diets [8, 9] Involvement towards creating the prevailing point-of-purchase surroundings a lot of causative to the choice of healthy choices [11, 13]. The employment of processers to research pictures has

several possible claims for mechanism skillful farmed everyday jobs. So as to boost the practicality and suppleness of the popularity scheme form and size options may be shared in conjunction by color and texture structures [12]. Results with SVM area unit clearly superior to those found with the easy Naïve Bayes classifier [14]. The model on several alternative object classes variable over compactness and form. In their case this refinement is answerable for a tenth increase in performance relative to their initial recognition hypotheses [15]. The scheme automatically discoveries the amount of each element familiarized make the nourishment oppression information providing on its tag composed with the nourishment data for a smallest of a number of the elements [16]. The typical for this duty may be useful to pre-processing of food article identification or cull ender the exploration outcomes of inquiries associated with food, meals or dishes. They achieved high accuracy ninety-six, ninety-five and ninety-nine within the 3 datasets respectively [17].

They cantered on characteristic food things in a picture by victimization image process and segmentation, food classification victimization SVM, food portion volume activity, and calorie activity supported food portion mass and biological process tables [18]. MT-Diet improves the accuracy of machine-controlled food identification to eighty-eight.93%, a twenty fifth increase with relation to competitive techniques [19]. A diet proof of identity structure that mixes international and native options for a lot of correct visual portrayal of food things. By smearing late call fusion-based rules to every discrete feature network we are able to growth the proper classification percentage by over 7% [20]. The processes of food recognition and calorie estimation area unit performed in cloud server. A mean recall rate of ninety.98%, preciseness rate of ninety-three.05%, and accuracy of ninety-four.11% compared to fifty.8% to half of 1-mile accuracy of alternative existing food recognition systems [21].

Mushrooms, in spite of the nice variability determined among species, represent a remarkable food item that may contribute to the formulation of a well-balanced diet [22]. The results of the current study showed that this modification may be wont to predict organic food shopping for intentions and self-reported shopping for behavior [23]. VBM could be a quick rising technology because of the increasing affordability and capability of camera and computing hardware/software, may be employed in recognition system [24]. The recovery of biotic process info directly from the recipes by oppression NLU/IE techniques. They used NLG. Techniques to supply the answer [25]. Algorithmic theorem learns from food intake of human. Improves accuracy for classification algorithm [26]. 79.2% classification percentage with the highest 5 applicants once ground-truth leaping containers square measure given. It's higher eleven.0% than methodology conveys available on a smartphone that accept color histogram [27]. 79.2% classification accuracy given and developed android based application [28].

A. Deep Learning

Deep learning is a module of a broader conception of machine learning ways supported learning knowledge and its illustrations. Learning are often carried in 3 ways managed, semi-supervised or unattended. Deep learning consists of following architectures like deep neural ne process, audio recognition, social network filtering, computational linguistics, bio-informatics, medical image analysis, material scrutiny and parlor game programs, wherever they need made results love and, in some cases, superior to human specialists.

B. Object Detection

Manatee recognition may be a skill associated with idea and image process that contracts with spying instances of similar objects of a particular category (such as Food, Animal, buildings, or cars) in digital pictures and videos. It is domains of item recognition signify face detection and ordinary finding. The applications of Object detection area unit employed in varied field of pc vision, as well as image reclamation and video police work. The categorification of each object class is be contingent on its own features and attribute

– For example, all Oranges are round. Object class detection uses these features and feature of every objects. For example, when looking at orange, it states its feature like shape, color, texture etc.

III. SYSTEM METHODOLOGY

Designed for the goal of object recognition maltreatment convolutional neural network, we have a inclination to probable a scheme that rounds on decent phones, and which document the employer to need an image of the nutrition and aware the amount of calorie feasting automatically. AN app that takes image of fruits and categorize object in different category and provides nutrition value once you snap an image of your plate, the app uses the pictures to form the comparison. It then provides you with an inventory of things that area unit the foremost doubtless foods. you will not see one choice with everything on your plate. Instead, you will see every known item listed separately.

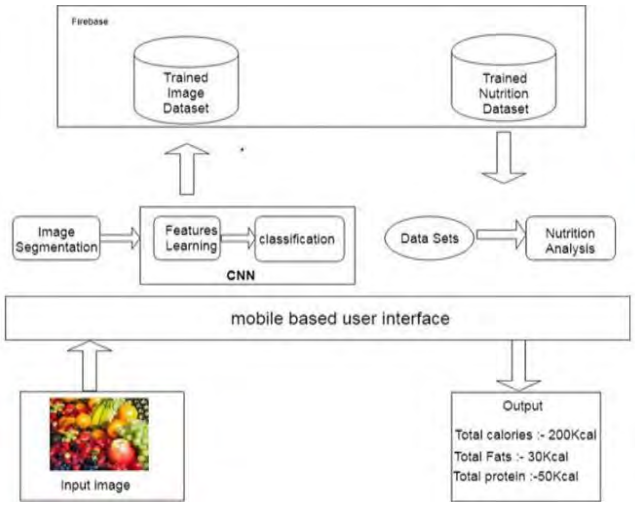


Fig.1. Proposed Architecture

This system is powered by Convolutional Neural Network. The method of object recognition is applied mistreatment Convolutional neural network mistreatment Tensor Flow. The system use base of operations cubic centimeter at the backend wherever the customaries trained model area unit host which can dynamically serve the user input. The output is generated by classifying the image and provides useful information to user such as Name, Calories & Diet. This technology is used to quickly and accurately identify food items.

This is an assistive calorie amount system to help patients and doctors embellishment in their match in contradiction of diet-related fitness situations. The discovery effectiveness can be achieved up to 90% for different foods captured from different position and surfaces. Every food item recognized by the system is paired with detailed nutrition information. Only if users/patients with suitable and smart resolutions that service them amount their nutrition consumption and gather dietary info in the direction of continuing anticipation and fruitful action plans. It's too benefit in the superstore and farming segment to stream well vegetal and fruits

A. *Basic CNN Components*

It is encouraged by biological procedures of neurons that resemble the connectivity outline amongst the neurons of the animal visual cortex. The specific cortical neurons reply to inducements only the rare area of the pictorial field. The pictorial arena of dissimilar neurons temperately overlays such that they cover the pictorial meadow. CNN is a scheme much like a multilayer perceptron that has been intended for shortened processing supplies. As connected to the other image organization algorithms, it has less pre-processing. CNN consists of a dissimilar layer which are an input layer, and output layer, and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers.

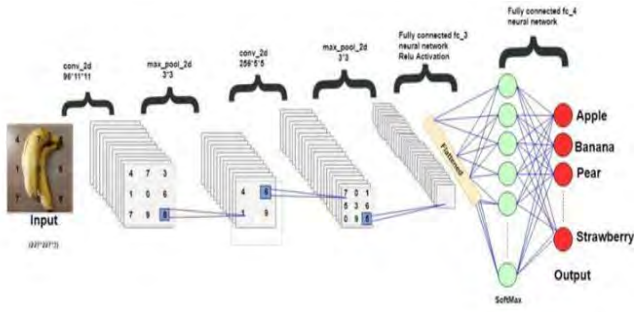


Fig. 2. CNN Architecture

The last hidden layer produces an output values forming a vector x =x. The output neural layer is can classify among n=1…n categories with a SoftMax activation function. Each of the n categories is assigned with conditional probabilities to classify. In each node in the output layer the pre-activated values will consist of the scalar products w(Tj)x where wj∈{w1,w2,…,wn}.However, the activation of this output layer will take through the softMax function which will map the vector of image to vector of n element in[0,1].

$$Zj = wTj * x$$

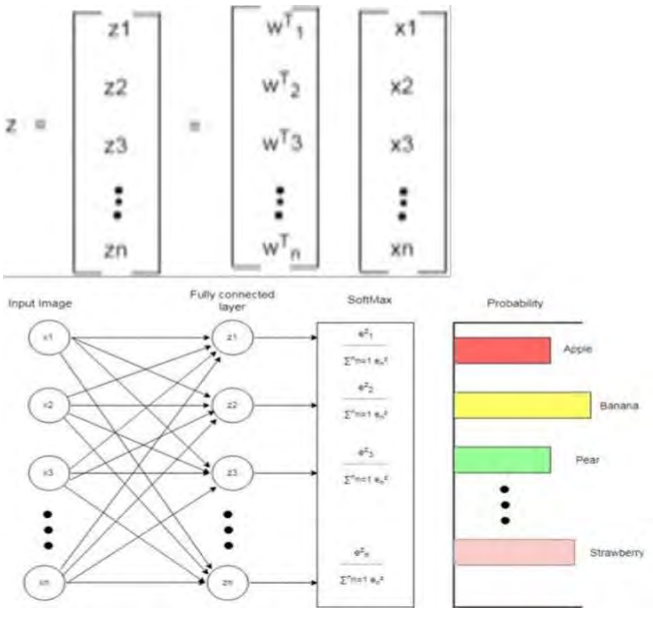


Fig.3. Soft Max Function

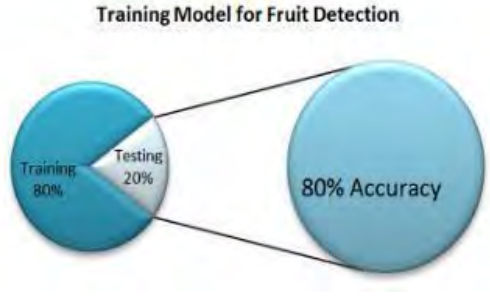


Fig.4. Training Model for Fruit Detection

In the Above figure, overall 1000 images of per object has been skilled in the exercise model for fruit detection amongst which 80% of data is skilled data and 20% of data is challenging data which delivers 80% of accuracy.

IV. EXPERIMENTAL CONDITIONS

Fig.5. CNN Result for object Detection for strawberry

In Fig.5 and Fig.6 following object are detected with their name and nutrition value of that object is shown after calculation which is done by Convolutional Neural Network (CNN).

Fig.6. CNN Result for object Detection of PEAR with Nutrition Values

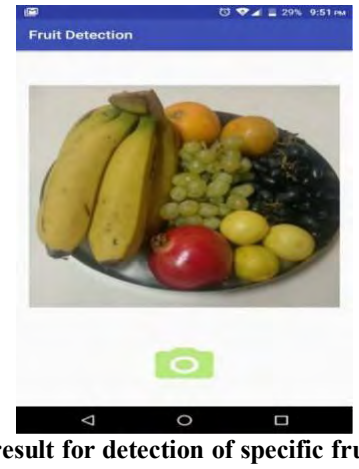


Fig.7. CNN result for detection of specific fruits from multiple food

In fig.7 object will be detected with their name after calculation which is done by Convolutional Neural Network (CNN).

V. CONCLUSION

In this paper, we have described the growth of an Object detection system using CNN, which run on mobile devices. We erected a object image dataset from capturing multiple images of a particular object, applied Convolutional Neural Network to the identification of 20 objects, and calculated its presentation. Convolutional Neural Network achieved much improved performance and efficiency than did old-style approaches using handcrafted structures. Complete comment of skilled convolution kernels, we inveterate that color geographies are important to food image recognition. We applied Convolutional Neural Network to object discovery, finding that Convolutional Neural Network expressively outperformed a baseline method. After recognition, algorithm fetches the object values of detected object and display it to the user.

Reference

1. Fruit Detection using Improved Multiple Features based Algorithm - Hetal
N. Pate, Dr. R.K.Jain, Dr. M.V.Joshi. International Journal of Computer
Applications (0975 – 8887) Volume 13– No.2, January 2011
2. Automated Grading of Palm Oil Fresh Fruit Bunches (FFB) using Neuro- Fuzzy Technique - Nursuriati Jamil, Azlinah Mohamed, Syazwani Abdullah. 2009 International Conference of Soft Computing and Pattern Recognition
3. On-tree fruit recognition using texture properties and color data - J. Zhao, J. Tow, J Katupitiya. IEEE Xplore: 05 December 2005
4. Online: Report on Fruit and vegetable classification from live video. - Lukas Danev. University of Edinburgh 2017
5. Automatic fruit recognition: A survey and new results using
Range/Attenuation images - A K jain, R Ceres, J L Pons. 7 June 2001.
6. H. Kagaya, K. Aizawa, and M. Ogawa, “Food Detection and Recognition Using Convolutional Neural Network,” Proceedings of the ACM International Conference on Multimedia - MM 14, 2014.
7. F. Zhu, M. Bosch, I. Woo, S. Kim, C. J. Boushey, D. S. Ebert, and E. J. Delp, “The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation,” IEEE Journal of Selected Topics in Signal Processing, vol.
4, no. 4, pp. 756–766, 2010.
8. W. Wu and J. Yang, “Fast food recognition from videos of eating for calorie estimation,” 2009 IEEE International Conference on Multimedia and Expo, 2009.
9. T. Joutou and K. Yanai, “A food image recognition system with Multiple Kernel Learning,” 2009 16th IEEE International Conference on Image Processing (ICIP), 2009.
10. V. Dewanto, X. Wu, K. K. Adom, and R. H. Liu, “Thermal Processing Enhances the Nutritional Value of Tomatoes by Increasing Total Antioxidant Activity,” Journal of Agricultural and Food Chemistry, vol.
50, no. 10, pp. 3010–3014, 2002.
11. .S. Arivazhagan1, R. N. Shebiah, S. S. Nidhyan, and L. Ganesan, “Fruit Recognition using Color and Texture Features,” Journal of Emerging Trends in Computing and Information Sciences, vol. 1, no. 2, Oct. 2010.
12. G. Cowburn and L. Stockley, “Consumer understanding and use of
nutrition labelling: a systematic review,” Public Health Nutrition, vol.
8, no. 01, 2005.
13. G. Csurka, C. Bray, C. R. Dance, and L. Fan, “Visual Categorization with Bags of Keypoints,” Workshop on Statistical Learning in Computer Vision,ECCV, pp. 1–22, 2004.

14. . R. Fergus, P. Perona, and A. Zisserman, “A Sparse Object Category Model for Efficient Learning and Exhaustive Recognition,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05).
15. P. Pouladzadeh, S. Shirmohammadi, and A. Yassine, “You are what you eat: So measure what you eat!,” IEEE Instrumentation & Measurement Magazine, vol. 19, no. 1, pp. 9–15, 2016.
16. . H. Kagaya and K. Aizawa, “Highly Accurate Food/Non-Food Image Classification Based on a Deep Convolutional Neural Network,” New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops Lecture Notes in Computer Science, pp. 350–357, 2015.

17. P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi, “Measuring Calorie and Nutrition From Food Image,” IEEE Transactions on
Instrumentation and Measurement, vol. 63, no. 8, pp. 1947–1956, 2014.
18. J. Lee, A. Banerjee, and S. K. S. Gupta, “MT-Diet: Automated smartphone based diet assessment with infrared images,” 2016 IEEE

International Conference on Pervasive Computing and Communications (PerCom), 2016.
19. M. Bosch, F. Zhu, N. Khanna, C. J. Boushey, and E. J. Delp, “Combining global and local features for food identification in dietary assessment,”

2011 18th IEEE International Conference on Image Processing, 2011.
20. . P. Pouladzadeh and S. Shirmohammadi, “A Cloud-Assisted Mobile Food Recognition System,” Electrical Engineering and Computer Science department, University of Ottawa, May 2017.

21. P. Manzi, L. Gambelli, S. Marconi, V. Vivanti, and L. Pizzoferrato,

“Nutrients in edible mushrooms: an inter-species comparative study,” Food Chemistry, vol. 65, no. 4, pp. 477–482, 1999.
22. . A. Tarkiainen and S. Sundqvist, “Subjective norms, attitudes and intentions of Finnish consumers in buying organic food,” British Food Journal, vol. 107, no. 11, pp. 808–822, 2005.
23. . S. Shirmohammadi and A. Ferrero, “Camera as the instrument: the rising trend of vision based measurement,” IEEE Instrumentation & Measurement Magazine, vol. 17, no. 3, pp. 41–47, 2014.
24. 24. A. Mazzei, L. Anselma, F. D. Michieli, A. Bolioli, M. Casu, J. Gerbrandy, and I. Lunardi, “Mobile Computing and Artificial Intelligence for Diet Management,” New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops Lecture Notes in Computer Science, pp. 342–349, 2015.

25. Y. Wang, Y. He, F. Zhu, C. Boushey, and E. Delp, “The Use of Temporal Information in Food Image Analysis,” New Trends in Image Analysis and

Processing -- ICIAP 2015 Workshops Lecture Notes in Computer Science, pp. 317–325, 2015.
26. Y. Kawano and K. Yanai, “Rapid Mobile Object Recognition Using Fisher Vector,” 2013 2nd IAPR Asian Conference on Pattern Recognition, 2013.

27. Y. Kawano and K. Yanai, “FoodCam: A Real-Time Mobile Food Recognition System Employing Fisher Vector,” MultiMedia Modeling Lecture Notes in Computer Science, pp. 369–373, 2014.

Comparing Techniques for Imputing Missing Values on Framingham Dataset

Shiwani Gupta, Karan Kukreja, AnkitJha, Divya Gupta
Computer Engineering,
Thakur College of Engineering and Technology
University of Mumbai
shiwanigupta@thakureducation.org, kukrejakaran45@gmail.com, jhaankit102@gmail.com, dipugupta2811@gmail.com

Abstract - **Data is the collection of various statistics which is fundamental for running the algorithms. The missing in the data can make the algorithms work differently. Hence, missing data is exciting but common probleminmachinelearningandstatisticalanalysis.They can be caused by various sources such as death ofpatients during ongoing treatment, users not responding in a survey, errors in the database during deletion or updating ofinformation,andsoon.Thekindofcategoriestheycome depends on the degree of relationship between the nature of the observed values and missing data. Thus, understanding the appropriate analysis for using the data needs the understanding of the mechanism for handling missingdata.There are various techniques methods have been used for handling the missing data in textbooks, literatures and standard courses. The importance of these techniques in medical science and clinical trials has made some of them gaining resurgence in the recent decade. But, there are also some drawbacks associated with some of them when used for data mining algorithms. Also, there is a need to avoid bias or the over- or under-estimation of variability.**

We are using here case deletion and some imputation techniques from various fantastic python packages like fancy-impute missingpy, sklearn-imputer and pandas. The techniques will be used to impute the missing values on Framingham dataset, which further will be examined for implementing different machine learning algorithms. This examination will help to analyze the best imputation technique by comparing them using different performance measures and parameters, such as accuracy, specificity, precision, ROC curve area, etc. The underlying principles of machine learning methods and model-based imputation methods can also be explored.

Keywords-- Missing, Data frame, fancy impute

I. INTRODUCTION

Thetechnologiesareevolvingwitheachday. Many of the technologies give out the results on the basisofdata.Oneofthecommonproblemsinmostof the scientific research domains such as medicine, Biology or Climatic Science is missing data. Measurement error, non-response or deleted aberrant value, low signal-to-noise ratio, and mishandling of samples are some of the common sources for missing data. It introduces an element of ambiguity into analysis of data. The missing in data can cause the propertiesofstatisticalestimatorssuchaspercentages, variances or means, resulting in misleading conclusionsandlossofefficiency.Thisproblemisthe most common issue encountered by the algorithmand machine learning practitioners when they analyze the real-world data. In many applications ranging from survey response in social sciences to gene expression in computational biology, various degree of missing data is present. As many machine learning algorithms and statistical models rely on whole data sets, it is necessary to handle the missing data appropriately. There are number of techniques have been proposed for imputing with statistical predictions in place of missingoremptytuples,theprocessisgenerallycalled as 'missing dataimputation'.

Thus, it is very important to have complete datasets for accurate use of the data and decision making processes. The many reasons for missing of datacanincludethemanualdataentryprocedure;data recorded and transferred errors, and incorrect measurement. One of the significant examples of real world data is clinical data as it plays an importantrole in discovering and implementing machine learning and data mining algorithms, the effect of clinical data on lifestyle can be serious, hence there is a need of manipulating and analyzing this type of data. , There are three types of missing values; (a) Missing at Random (MAR), and (b) Missing Not at Random (MNAR) (c) Missing Completely at Random (MCAR). The categorization has been done on the relation between the incomplete attributes and other completedattributes.Whenthemissingvaluedoesnot depend on any of incomplete variables or other completed variables, the missing value comes under MCAR. It means the attribute A being missing is independentoneithercompletedataBorA.Whenthe variable being missing is independent from the values

of that variable, but it somehow depends on other dataset variables, i.e. where the distribution ismissing itselfbutdependonothervariable'sdata,suchmissing data is MAR. The last category, MNAR is non-ignorablecasewhichisneitherMCARnorMAR.The probability of data being missing from an attribute depends on the value of that attributeitself.

II. LITERATURE SURVEY

In the papers we have studied, the different imputation methods were chosen to cover techniques broadly applied in the literature and representative of various statistical strategies

Briefly, three out of these methods were based on mean, K-Nearest Neighbour (KNN), Fuzzy K-Means (FKM) while other were Singular Value Decomposition(SVD),BayesianPrincipalcomponent Analysis (bPCA) and Multiple Imputations by Chained Equations (MICE). The imputation using mean consists of substituting the missing value for a givenvariablebytheaverageoftherowvaluesofthat variable; KNN is the K-Nearest Neighbour where it substitutes the each missing value by average of the individual sample or set of K-nearest neighbours. FKMisFuzzyK-meanswhichisanextensionofKNN based on clustering. SVD is Singular Value Decomposition which imputes value based on the Eigen values while bPCA is a Bayesian Principal component Analysis which again uses Eigenvaluesfor imputing the missing value. Finally, MICE are an iterative algorithm based on chained equations which usesimputationmodelspecifiedseparatelyforeachof the variable while involving other variables as predictors.[5] There are some numerous modelsbased imputation techniques that can be categorizedinto explicit model based and implicit model based. Expected Maximization (EM) is one of the examples of explicit model while implicit model includes methods like K-nearest neighbor (KNN), weight KNN, RBFN and Multi-Layer Perceptron (MLP). Another way of categorization of imputation is single imputationandmultipleimputation.Singleimputation is simple method like mean imputation or Expected Maximization (EM). Multiple imputation include imputationmultipletimesonsamemissingvaluesuch asMICE[3].Theobjectiveoftheresearchreportedin previous papers is to manipulate and compare a set of imputation methods on different size of datasets like small datasets, large datasets and analyze them over different parameters like RMSE, UCE, SCE, and Execution Time. This analysis over these parameters will result the most useful imputation method to impute the missing value datasets[1].

Googlesearchofanymethodassociatedwith term 'missing data' gives an idea of their respective popularity. As expected, Mean has produced the most number of hits approximately 210000 results, followedbyMICE,SVDandKNN(17700,14500and 12700respectively.bPCAandFKMhasgiventheless popular around 2560 and 5220 hits respectively. However,thebestmethoddoesnotneedtobepopular. So, there was need of some performance measuresfor comparing these methods. These performance measures are: RMSE, UCE, SCE and the execution time[2].Theresultsweremeasureddifferentforsmall datasets and larger datasets. The overall summary is given in the table below. For the small datasets, FKM has given better RMSE, UCE and SCE than other methods but Execution time was not good. While for large datasets, bPCA has been best for all the performance measures. Thus, overall bPCA andFKM has outperformed the other methods. KNN, SVD and MICE has also given good results. Mean method has not been so good for imputation purpose[4].

III. METHODOLOGY FOR IMPUTATION

Computers has provided with the advance ability to learn without being explicitly programmed. This advanced ability is a Machine Learning which is a type of Artificial Intelligence. Machine learning

helps to develop the computer programs that can change when exposed to new data. The Python community has developed many modules to help the programmers implement machine learning. The many machine learning packages like fancy impute, missingpy, datacompy are installed. The many machine learning libraries like pandas, numpy, scikit-learnhasbeenusedtoimplementthemachinelearning models.

DROPPING ROWS WITH MISSING VALUES:

One of the simplest strategies for handling missing is to eliminate records that contain a missing value.Thatmeanstherowsinthedatasetthatcontains

Steps for Dropping Columns:

Step 1: Analyze Data Frame with NaN Values

Step 2: Drop the Columns with NaN Values in Pandas Data Frame

Step 3 (Optional): Reset the Index

IMPUTATION BY REPLACING WITH MEAN

VALUE:

The imputation is the technique to using a model to replace missing values. There are differenteven of the missing value tupple, that row will be dropped from the dataset. Pandas provides the dropna() function which can be use to remove all the rows from the dataset that contains the missing data. The easiest way to dropping rows is to remove rows containing NA; either when any column contains NA or all columns contain NA. Dropping the rows containing NA values in any of the column will result in reduction in number of rows. This effect can cause the datasets with less number of rows where dropping the rows with any missing value can cost us losing of necessary information.

Steps for Dropping Rows:

Step 1: Analyze Data Frame with NaN Values

Step 2: Drop the Rows with NaN Values in Pandas Data Frame

Step 3 (Optional): Reset the Index

DROPPING COLUMNS WITH MISSING VALUES:

Apart from dropping the rows, another strategyforhandlingmissingdataistodropthewhole column that contains the missing value tupple. This can be done by removing all the columns that contain any of the missing value. Pandas provides the data.drop () function where data() can be used to remove the columns. But the problem becomes significantly more complicated, when working with the data frame because the degree of information loss issohigh.Itmayalsohappenthatthecolumnoftarget variable contains the missing value and gets dropped by applying this method. If the column is sparsely populated, it might drop the columninstead.
ways which can be considered when replacing amissing value. When the data is in continuous form, replacingwiththemeanvaluecangivethebestresults. The imputing performed on training data will have to beperformedinnewdatawhenpredictionsareusedto finalize the model. There is need of consideration when selecting the way of imputation of missing values. Imputing the missing values with mean is just substitutingthemissingvaluesbyaveragevalueofthe complete column. All the missing values in that particular column will be the same evaluated average value. Pandas provide fillna() function for imputing the mean value in place of missingvalues.

IMPUTATION BY REPLACING WITH MODE VALUE:

One of the other imputation methods is by substituting the most frequent value of that column in place of missing values. This imputation method is called as Mode imputation method. Mode is the most repeated value of the particular column. This method does not involve the evaluation of any value but only toselectthevaluewhichgetsrepeatedmostnumberof timesinthecolumn.Thiscanbedonebytheusingthe pandas library which provide fillna() function for imputing the mean value in place of missing values. For using the mode method by fillna() function, the type used will be 'mode'.

Steps for Replacing with Mode Technique:

Step 1: Analyze Data Frame with NaN Values

Step 2: Compute the Mode of the column containing the NaN value

Step 3: Fill the missing tupple with the computed Mode Value using fillna () function

Step 4: Repeat the process for another column containing NaN value

IMPUTATION USING KNN ALGORITHM:

K- Nearest Neighbour is an algorithm which is useful for matching a point with its K closest neighbours in a multi-dimensional space. It can be used for many kind of data like discrete, continuous, categorical and ordinal which makes it particularly useful for dealing with all kind of missing values.
The strategy behind using KNN imputation method is that a point value can be approximated by thevaluesofitsclosestpoints,basedonothervariables. Thereneedtoconsiderationofmanyparameterswhen using KNN; one is to look for number of neighbours (value of K). If the value of K is low, it will increase theinfluenceofnoiseandtheresultsgeneratedwillbe less generalizable. While high value of K will tend to blur local effects which we are actually looking for. Thus it is recommended to take an odd value of K for binary classes to avoidties.

Steps for Replacing with Mean Technique:

Step 1: Analyze Data Frame with NaN Values

Step 2: Compute the Mean of the column containing the NaN value

Step 3: Fill the missing tupple with the computed Mean Value

Step 4: Repeat the process for another column containing NaN value

KNN Algorithm:

Step 1: Analyze Data Frame with NaN Values

Step 2: Select the number of neighbours i.e. value of K

Step 3: Compute the Euclidean Distance of empty tupple with its neighbour

Step 3: Fill the missing tupple with the computed Value using KNN imputer

IMPUTATION USING MULTIPLE IMPUTATIONSBYCHAINEQUATION(MICE) ALGORITHM:

same record with some the multiple iterations, the process is called multiple imputations. A imputation technique based on the multiple imputations is called asMultipleImputationbyChainedEquations(MICE). This technique has lot of advantages over the single imputation. MICE technique work with the assumption that the kind of missing data is missing at Random (MAR) data. The strategy behind workingof this technique is to run the multiple regressionmodels and each missing value is modelled conditionally depending on the observed (non-missing) values. MICE technique uses all the other variables in the dataset to predict any other missing value. Fancy imputer library provides special function mice that willdetectwhichvariablearethedatasethavemissing information. There is a matrix called as predictor matrix which gives information about the variables in the dataset were used to produce predicted values for matching. Consider an example where variables x1,x5,y2-y5 were used to create predicted values for y1.As we did not mentioned any seed value, the algorithm by default chose one random value; so if thereisanyneedforreproducingtheimputation,there could set a seed value for random numbergenerator.

MICE Algorithm:

Step 1: Analyze Data Frame with NaN Values

Step 2: Predict the value using the nearby variables in the Data frame for the tupple containing value NaN

Step 3: Converge the value to the nearest value of prediction

Step 4: If the stopping condition is met

IMPUTATION USING MISS FOREST ALGORITHM:

Random forest algorithm is one of the efficient machine learning algorithms for implementing the machine learning models. There is an extended version of Random Forest algorithm which can be used for imputing the missing value in the dataset. This extended imputation technique is knownasMissForestalgorithm.Bydefault,themodel

On the basis on number of times of imputation can be done, it can be distributed into two categories; single imputation and multiple imputations. When the value is getting imputed to beginstheimputationofmissingvaluesofthecolumn (which is expected to be a variable) with the least number of missing values (let his column be called as candidatecolumn).Theinitialstepistofillinganymissing values of the non-candidate, remaining, columns with an initial guess, which is the column mode for all the columns representing categorical variables while column mean representing for numericalvariables.Thecategoricalvariablesmustbe explicitly identified during the imputer's fit () method call. After the initial step, the imputer fits a random forest model with the candidate column as output variable and the remaining columns as the predictors overallrowswherethecandidatecolumnvariablesdo nothaveanymissingvalues.Afterthefittingofmodel, the rows with missing values of candidate column are imputed using the prediction form fitted Random Forest. The non-candidate column rows act as input dataforfittedmodel.Further,theimputermovesonto next candidate column with the second smallest number of missing data among the non-candidate columns in the first round. The whole process repeats itself for each of the column having missing values, possibly over epoch or multiple iterations for each column, until the stopping condition ismet.

Miss Forest Algorithm:

Step 1: Analyze Data Frame with NaN Values

Step 2: Compute Normalized Root Mean Squared Error for the tupple containing value NaN

Step 3: Iterate the steps until stopping Condition is met

Step 4: If the stopping condition is met for iteration

Do: Impute the computed value in empty tupple

Step 5: Repeat the process for another tupple in Data frame containing NaN value

METHODOLOGY FORPREDICTION

LOGISTIC REGRESSION

A statistical method for analyzing thedataset having one or more independent variable that determines an outcome which is measure is with a dichotomous variable. This method is known as Logistic Regression. In this regression model, the target variable is dichotomous variable i.e. target variable only contain data codes as 0 (false/failure) or 1 (true/success).

DECISION TREE

A learning method for approximation of discrete-valued target functions and the function is represented by set of if-then rules in the form of tree shape structure is known as Decision Tree Learning method. These learned trees act as model to fit the independent variables data as conditions and outcome variable as decisions based on the respective conditions.

RANDOM FOREST

Decision Tree model used for classification can give efficient results. Another model can also be used for classification to enhance the efficiency of results. Instead of using a single decision tree model, we can collective use group of decision tree models whichcanproducetheresultslikeforest(collectionof trees). This algorithm is called as Random Forest algorithm and the classifier used by Random forest is called as Random forest classifier. In the random forest, each individual decision tree spits out a class prediction and the class which consists of most number of votes turn out to be the model'sprediction.

METHODOLOGY FOR EVALUATING PERFORMANCE

- ACCURACY
- SENSITIVITY
- SPECIFICITY
- PRECISION
- F1-SCORE
- ROC-AREA UNDERCURVE

SYSTEM ARCHITECTURE

The architecture of system consists of some finite steps. The process begins with large dataset without any missing value in the dataset. After that, thedatasetisintroducedwithsomeamountofmissing value by deleting the random values of records of randomly selected rows or columns. The degree of missing value introduced can be ranged from 5% to 45%. If the dataset which is already incomplete and percentageofmissingvaluesmorethan45%,thenuse ofsuchdatasetsshouldbeavoided.Thenextstepisto apply the imputation techniques on the dataset. The datasets which are made complete by imputing the missing values by imputations techniques are applied by some machine learning algorithms. The results generatedarethencomparedonthebasisofsomeperforman ce measures parameters. The results are analyzed for multiple times of simulation and the conclusions are made about the best imputation technique.

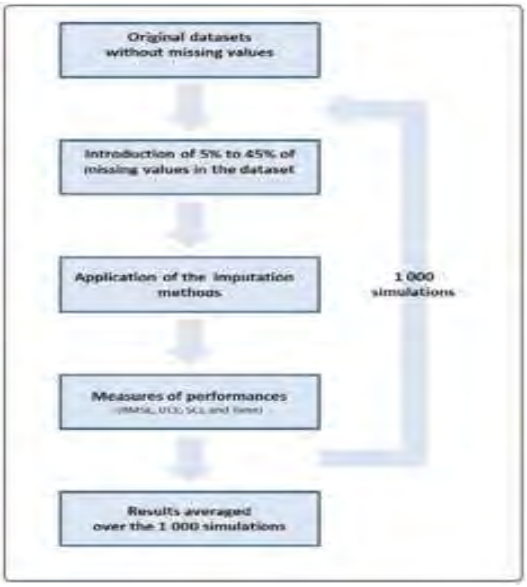


Fig 1. Basic flowchart for Imputation process using Algorithms

IV. RESULTS AND DISCUSSIONS ACCURACY

Table I. Accuracy of the model fitted to datasets by

ALGORITHMS / IMPUTATION TECHNIQUES	LOGISTICS REGRESSION	DECISION TREE	RANDOM FOREST
DROPPING ROW	0.8561	0.7486	0.8251
DROPPING COLUMN	0.8498	0.7517	0.8317
MEAN IMPUTATION	0.8537	0.7507	0.8356
MODE IMPUTATION	0.8538	0.7421	0.8427
KNN IMPUTAION	0.8537	0.7515	0.8388
MICE IMPUTATION	0.8530	0.7445	0.8317
MISS FOREST IMPUTATION	0.8537	0.7531	0.8396

The above table is depicting the accuracy values of the model obtained from the results of each of the algorithm fitted to the Framingham dataset whose missing values are imputed by the aboveseven different imputation techniques. The highest accuracy

is 0.8561 for the model Logistics Regression when fitted to dataset imputed by dropping row imputation technique. For the Decision Tree model, the maximumaccuracyobtainedis0.7531forMissForest Technique while the same gives accuracy of 0.8396 which is highest for Random forestmodel.

SENSITIVITY

Table II. Sensitivity of the model fitted to datasets by different imputation techniques

ALGORITHMS / IMPUTATION TECHNIQUES	LOGISTICS REGRESSION	DECISION TREE	RANDOM FOREST
DROPPING ROW	0.8574	0.8532	0.8415
DROPPING COLUMN	0.8419	0.8564	0.8544
MEAN IMPUTATION	0.8542	0.8597	0.8504
MODE IMPUTATION	0.8543	0.8596	0.8560
KNN IMPUTAION	0.8542	0.8591	0.8509
MICE IMPUTATION	0.8536	0.8545	0.8492
MISS FOREST IMPUTATION	0.8537	0.8560	0.8538

The above table is for Sensitivity values of the model obtained from the results of each of the algorithm fitted to the Framingham dataset whose missing values are imputed by the above seven different imputation techniques. The highest sensitivityvalueis0.8597forthemodelDecisionTree when fitted to dataset imputed by Mean imputation technique. For the Logistics Regression model, the maximum sensitivity obtained is 0.8574 for Dropping Row method while the Random forest model gives maximum sensitivity of 0.8360 when fitted to dataset imputed by Mode Imputationtechnique.

SPECIFICITY

Table III. Specificity of the model fitted to datasets by different imputation techniques

ALGORITHMS / IMPUTATION TECHNIQUES	LOGISTICS REGRESSION	DECISION TREE	RANDOM FOREST
DROPPING ROW	0.7272	0.2552	0.3684
DROPPING COLUMN	0.5555	0.2087	0.3392
MEAN IMPUTATION	0.8125	0.2211	0.3142
MODE IMPUTATION	0.8126	0.2155	0.4523
KNN IMPUTAION	0.8125	0.2196	0.3548
MICE IMPUTATION	0.8000	0.1971	0.2631
MISS FOREST IMPUTATION	0.8571	0.2079	0.4000

The above table is for Specificity values of the model obtained from the results of each of the algorithm fitted to the Framingham dataset whose missing values are imputed by the above seven different imputation techniques. The highest Specificity value is 0.8571 for the model Logistics Regression when fitted to dataset imputed by Miss Forest imputation technique. For the Decision Tree model, the maximum Specificity obtained is 0.2211 for Dropping Column method while the Random forest model gives maximum Specificity of 0.4523when fitted to dataset imputed by Mode Imputation technique.

PRECISION

Table IV. Precision of the model fitted to datasets by different imputation techniques

ALGORITHMS / IMPUTATION TECHNIQUES	LOGISTICS REGRESSION	DECISION TREE	RANDOM FOREST
DROPPING ROW	0.9967	0.8449	0.9803
DROPPING COLUMN	0.9962	0.8522	0.9702
MEAN IMPUTATION	0.9972	0.8520	0.9879
MODE IMPUTATION	0.9973	0.8327	0.9739
KNN IMPUTAION	0.9973	0.8447	0.9786
MICE IMPUTATION	0.9972	0.8457	0.9842
MISS FOREST IMPUTATION	0.9981	0.8382	0.9804

The above table is for precision values of the model obtained from the results of each of the algorithm fitted to the Framingham dataset whose missing values are imputed by the above seven different imputation techniques. The highestprecision value is 0.9981 for the model Logistics Regression when fitted to dataset imputed by Miss Forest imputation technique. For the Decision Tree model, the maximum precision obtained is 0.8522 for Dropping Column method while the Random forest modelgivesmaximumprecisionof0.9879whenfitted to dataset imputed by Mean Imputationtechnique.

F1-SCORE

Table V. F1-Score of the model fitted to datasets by different imputation techniques

ALGORITHMS / IMPUTATION TECHNIQUES	LOGISTICS REGRESSION	DECISION TREE	RANDOM FOREST
DROPPING ROW	0.9218	0.8482	0.9034
DROPPING COLUMN	0.9184	0.8526	0.9102
MEAN IMPUTATION	0.9202	0.8491	0.9155
MODE IMPUTATION	0.9203	0.8460	0.9109
KNN IMPUTAION	0.9202	0.8555	0.9105
MICE IMPUTATION	0.9198	0.8492	0.9149
MISS FOREST IMPUTATION	0.9203	0.8481	0.9146

The above table is for F1-Score of the model obtained from the results of each of the algorithm fittedtotheFraminghamdatasetwhosemissingvalues are imputed by the above seven different imputation techniques. The highest F1-Score is 0.9218 for the model Logistics Regression when fitted to dataset imputed by Dropping Row imputation technique. For the Decision Tree model, the maximumF1-Scoreobtained is 0.8555 for KNN imputation method while theRandomforestmodelgivesmaximumF1-Scoreof 0.9155 when fitted to dataset imputed by Mean Imputationtechnique.

ROC-AREA UNDER CURVE

Table VI. ROC-Area under Curve of the model fitted to datasets by different imputation techniques

ALGORITHMS / IMPUTATION TECHNIQUES	LOGISTICS REGRESSION	DECISION TREE	RANDOM FOREST
DROPPING ROW	0.5229	0.5516	0.5121
DROPPING COLUMN	0.5111	0.5230	0.5411
MEAN IMPUTATION	0.5314	0.5332	0.5271
MODE IMPUTATION	0.5316	0.5439	0.5354
KNN IMPUTAION	0.5317	0.5652	0.5199
MICE IMPUTATION	0.5292	0.5223	0.5329
MISS FOREST IMPUTATION	0.5296	0.5390	0.5412

The above table is for ROC-Area under Curve (AUC) of the model obtained from the results of each of the algorithm fitted to the Framingham dataset whose missing values are imputed by the above seven different imputation techniques. The largest AUC is 0.5652 for the model Decision Tree when fitted to dataset imputed by KNN imputation technique. For the Logistics Regression model, the maximum AUC obtained is 0.5317 for KNN imputation method simultaneously while the Random forest model gives maximum AUC of 0.5412 when fitted to dataset imputed by Miss Forest Imputation technique.

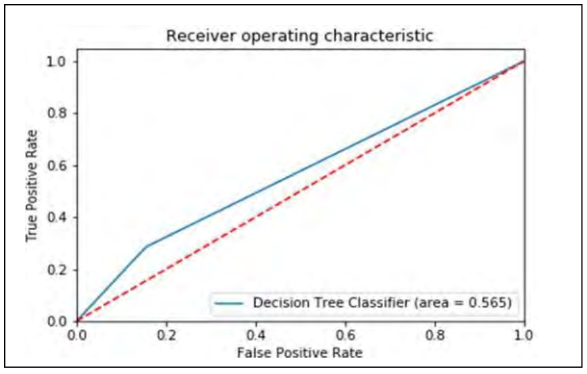


Fig 2. ROC curve for Decision Tree Classifier fitted on dataset imputed by KNN Imputation Technique

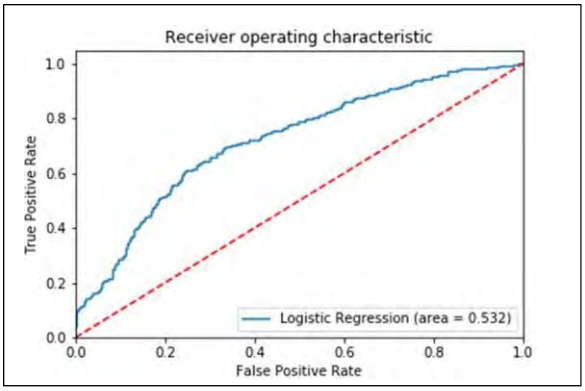


Fig 3. ROC curve for Logistic Regression fitted on dataset imputed by KNN Imputation Technique

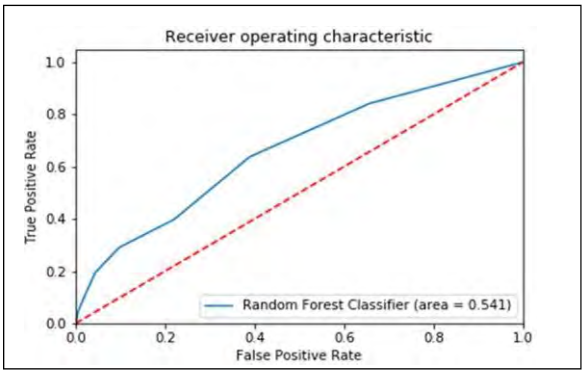


Fig 4. ROC curve for Random Forest fitted on dataset imputed by Miss Forest Imputation Technique

V. CONCLUSION AND FUTURE SCOPE

The complete datasets generated used for fitting on the three different machine learning models viz. Logistic Regression model, Decision Tree Classifier and Random Forest classifier. Several resultsobtainedarethenmadecomparisononthebasis of six different performance measures i.e. Accuracy, Sensitivity,Specificity,Precision,F1-ScoreandROC-Area under curve. The simple imputations like Dropping Row and Dropping Column led the loss of information while Mean and Mode imputations filled the same value for a column. The MICE and Miss Forest imputation methods involved lot of evaluation of values to impute. Thus, the ROC-Area under curve ofDecisionTreeislargestwhenappliedonthedataset imputed by KNN imputation technique. That means Decision Tree algorithm works well among the three algorithms. The ROC curve area obtained on dataset imputed by KNN gives largest area for two different machine learning models. Therefore the Framingham dataset imputed by KNN imputation imputes best

values among the other imputation techniques. Also KNN imputation method over other is that itimputeswithout loss of information. Further, the complex cases like imputing for nominal or longitudinal data can also be considered for analyzing in future with some careful attention. There is also some area of improvement where the advance performance measures like Cohen-Kappa score could be used for better understanding and comparison of the models fitted to imputed datasets.

References

[1] Mohammad Al Khaldy and ChandrashekharKambhaampati" Performance Analysis Of Various Missing Value Imputations Method On Heart Failure Dataset" Volume 118, No. 24, May 2018, International Journal of Pure and Applied Mathematics, Sinhgad Institute of Technology, Lonavala,Maharashtra,India.

[2] Schmitt P, Mandel J, Guedj "A Comparison of Six Methods for Missing Data Imputation" IEEE transactions on learning technologies, vol. 3, no. 3, July-September2016.

[3] AtulKumarPandey,PrabhatPandey,K.L.Jaiswal, Ashish Kumar Sen "DataMining Clustering Techniques in the Prediction of Heart Disease using Attribute Selection Method" International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 10, October2013.

[4] Syed Muhammad Saqlain , Muhammad Sher,Faiz Ali Shah, Imran Khan, Muhammad Usman Ashraf "Fisher Score And Matthews Correlation Coefficient- Based Feature Subset Selection For Heart Disease Diagnosis Using Support Vector Machines"Springer- Verlag London Ltd., 14 March2018.

[5] K. Z. Mao "Feature Subset Selection For Support Vector Machines Through Discriminative Function Pruning Analysis" Ieee Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, Vol.34, No. 1, February2004

[6] http://www.machinelearningmastery/impuationtechniques.Html(23/09/2019).

[7] http://www.towardsdatascience/missingvalueimputations.Html(29/10/2019).

[8] http://www.analyticsvidhya/sixtechniquesforimputation.Html(30/10/2019).

[9] http://kaggle/framinghamdataset.Html(30/10/2019).

[10] https://medium.com/handlng-missing-values-in- machine-learning-part(07/11/2019)

Design and Development of System for Music Instruments Identification Using SVM

Pooja Trivedi,Suraj Maurya,Pratik Baheti,Vaishali Nirgude
Thakur College of Engineering and Technology
Mumbai, India
surajmaurya029@gmail.com,pratikbaheti13@gmail.com,vaishali.nirgude@thakur education.org

Abstract— **Music instruments are hard to classify as many of them have a very minor difference in their sound signature. It is hard even for humans a while to differentiate between similar sounding instruments. Main focus of this system is to classify instrument recording. This identification system has been implemented by various algorithms like K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN), Naïve Bayes, Support Vector Machine (SVM), etc. By comparing the algorithms, SVM gave better accuracy than K-means, Naïve Bayes and it is also less complex than CNN. Therefore, the proposed system will be implemented by using SVM. All the previous work is done on static dataset only but this system is also focusing on dynamic data as well. The input to this model consist of individual audio recording for each instrument in Wav format. To perform feature abstraction firstly audio file is processed by the district Fourier transform to normalized signal and label each example with a unique integer corresponding to instrument class. Then by MFCC technique features are extracted and use this data to classify, analyses and train multi- class SVMs to make predictions on the test set. The system is expected to give approximately more than 90% of accuracy as outcome.**

Keywords— **Music Instrument Identification, SVM, Prediction of Musical Instrument, Classification of Instruments, MFCC.**

I. INTRODUCTION

Music is formed by using various instruments and vocals from a person's in most cases. Music instruments are hard to distinguish as they have a very small difference in their sound architect. Even humans cannot properly differentiate between similar sounding instruments. Instrument sounds played by Different artists are also different as each one of them have their own style and character and also the quality of the instrument brings some differences. Recognizing which instrument is playing in a music will help in recommending songs to a user more accurately as by knowing persons listening habit we can know if they prefer a particular instrument and make better recommendations to them. These suggestions could be tailored according to a person's liking. Searching for music will also be benefited as from the longest time only way songs or music has been searched is by typing its name or an instrument, this can be optimized as we can add more filters to the search and fine tune the parameters instruments. Anyone who is searching for an artist can easily sort songs based on instruments which are being played in the background. It can also help to see which instruments are more popular and where this may help to cater more precisely to a selected portion of the populous [1]. Using SVM to classify images has been extensively researched and has produced highly accurate results. Using these detailed results and applying them to segregate instruments is our goal. Ample research has also been carried out to classify instruments. SVM have been used for classifications in many applications since many years with great results and to apply these to instrument recognition in music is the key idea which may help to achieve satisfactory results.

II. LITERATURE SURVE

TABLE I

ANALYSIS OF LITERATURE SURVEY

Sr. No	Title	Author, Year and Publication	Description	Advantages	Gap Identified/ Future Scope
1.	Musical Recognition CNN and SVM.[1]	Prabhjyota Singh, Dnyaneshw Bachhav, Omkar	Audio excerpts is Preprocessed into	CNN and SVM result in Higher Accuracy.	CNN is complex.

		Joshi, Nita Patil, 2019(IRJET).			
2.	Musical Identification Status Finding MFCC.[2]	Miss. Vichare, Prof. S. Gulhane, (IJERD)	K-NN& SVM using feature vector generated.	MFCC is good for extracting features.	Audio features such as wavelet transform can be used.
3.	Learning Identification.[3]	Lewis Guignard and Greg Kohoe, 2015	Features Extracted using DFT. PCA, K-means and SVM were applied.	SVM better than PCA and K-Means	Focused only on 8 instruments commonly used in rock band.
4.	SVM-Based Automatic Classification of Musical Instruments.[4]	Jing Liu, Lingyun Xie, 2010 (IEEE).	Classification of Chinese and western instruments using SVM.	SVM properly classify the Instruments.	To implement for real Systems for MIR.
5.	Musical Sound with Deep using Feature Approach.[5]	Taejin Park Taejin Lee (ETRI).	MRPs with information is also with conventional with a multi-Network.	Using this Approach performance improved over system that uses a spectrogram.	Deep learning this fusion is complex requires domain Experience.
6.	Music using SVM.[6]	Ms. R. Makandar	Sequential Optimization algorithm Used for training SVM.	SVM shows better performance in music classification.	Can implemented Real time systems
7.	Musical Recognition Techniques, International of IT.[7]	A.S. Patankar, 2017	Wavelet Transform is used for feature extraction.	Study of Classificatio of Different	-
8.	A Study of Instrument Classification Gaussian Models and SVM.[8]	Janet Marques Pedro J. 1999	Features examined linear coefficients, FFT cepstral coefficients, FFT based mel cepstral coefficients.	Using Mixture and Accuracy was good.	Error rate was 30%.
9.	Musical Recognition Pairwise Classification	Slim Essid, Richard, IEEE, and David, 2006 (IEEE).	Inertia Ratio with feature projection and algorithms are considered	Higher rates can be with Optimized features.	Better when radial function Used.

	Strategies.[9]		using GMMs and SVMs.		
10.	Musical Sound Classification.[10]	R Rayar, M Anto Bennet, A Nazreen Banu, A Sushanthi, M	PSVM is trained and tested using DWT features.	Shows an accuracy of approximate 89%.	This can be for Real Applications MIR.

III. PROPOSED WORK

In this model, the dataset is applied on different algorithms like KNN, K-means, SVM, etc. and because of the accuracy measure of SVM, it will be being used for implementation of this system.

Support Vector Machines (SVMs) are one among foremost popular strong machine learning algorithms for classification, for their simple use and wide success. SVM is one among the efficient algorithm in machine learning. It is systematic in computation and vigorous in high dimension. SVM can also be used for classification as well as regression problems. The architecture is very much simple. The number of features required for recognition is less.

All the previous work is implemented on the static dataset. So in this model we are trying to focus on Dynamic dataset as well. The basic block diagram of the system is, as shown in the Fig. 1.

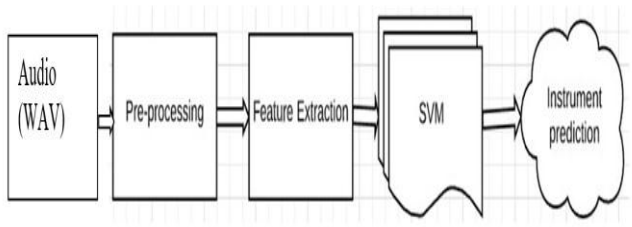


Fig. 1: Block Diagram

A. Dataset

The database is maintained which consists of all the different sound samples of musical instruments. The Database of music are the sets which have been classified into two sets i.e. training and testing.

B. Pre-Processing

Then in preprocessing step, it focuses on two stages of reading sounds and noise removal. To increase the accuracy of instrument recognition and to improve the strength of the signal Noise Reduction is used. Pre-processing is also used because the efficiency of subsequent feature extraction and classification stages can be increased so that it can enhance the overall retrieval performance.

C. Feature Extraction

After completion of pre-processing, main part of the project is executed- feature extraction. For different musical instruments different features will be extracted. Feature extraction is used to minimize the amount of data and to extract the meaningful and useful information from the given audio signal for a particular retrieval task.

D. Applying Algorithm

Further in the next step SVM algorithm is applied on the dataset to classify the instruments. Support Vector Machine works on principle based on training through inputs, using supervised learning techniques to classify data. This algorithm have somewhat complex hierarchy and design but it provides good results. The biggest plus point of SVM is that it easy to train. After the dataset passes through the algorithm, the result is then generated which consist of predicted instruments.

IV. METHODOLOGY

In ML, support vector machines (Support Vector Networks) are supervised learning models with associated learning algorithms that analyze the processed information and recognize patterns which are used for classification and multivariate analysis [5]. Given a group of training examples, each clear as belonging to at least one of two categories, an SVM training algorithm builds a model that allots new examples into one category or the opposite, making it a non-Probabilistic binary linear classifier. An SVM model may be a Representation of the examples as points in space, mapped in order that the samples of the separate categories are divided by a transparent space that is as large as possible [3].

SVM is a useful technique for data classification. Even though it's considered that Neural Networks are easier to use than this, however, every so often unsatisfactory results are obtained. A classification task usually involves with training and testing data which contains some data instances. Each example in the training set contains one target values and several attributes. The objective of SVM is to generate a model which predicts target value of knowledge instances within the testing set which are given only the attributes.

The Fig. 2 shows the process of Feature Extraction.

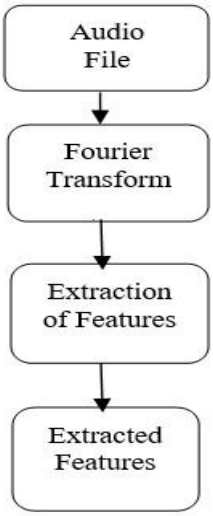


Fig. 2: Feature Extraction

Firstly, the audio file that is taken for consideration.

Then the audio fie is processed by using the Fourier Transform. It consist of transforming the sample of audio from time domain to frequency domain. This process results int frequency spectrum.

Feature extraction helps in representing the examples to be classified in terms of feature vectors or pairwise similarities. In this system, MFCC is used for extraction of features.

Finally, the extracted features are used further for classification of music instruments.

MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

MFCC is one of the standard feature among all feature set. MFCCs are consider as a way of representing the spectral information in a sound. Each and every coefficient has a value for each frame of the sound. MFCC is now being utilized in speech research and voice identification systems. In the sound identification system, the Mel-scale frequency is analogous to the perceived frequency, and is the most ordinary used simple frequency scale transformation equation, as described below. Let Mel be the Mel frequency and f be the actual frequency [4].

$$mel(f) = 2595 \cdot \log_{10}(1 + \frac{f}{700})$$

Fig. 3: MFCC Formula

The following Fig. 4 shows the basic working of SVM [6].

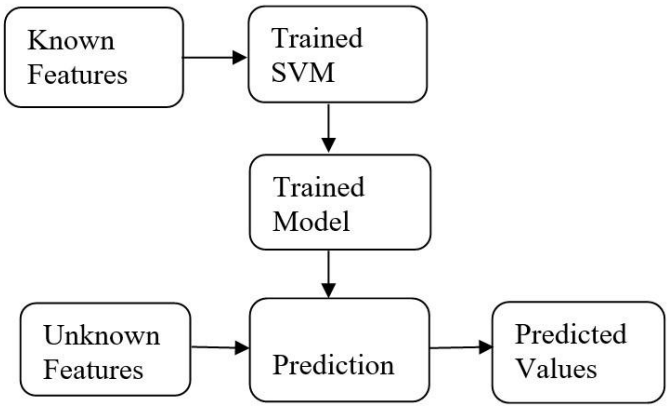


Fig. 4: Prediction of values using SVM

The Algorithms works on two datasets- training and testing.

The features of known are the training set and are passed for training the model. If the algorithm is trained properly by using sufficient training dataset, then it will predict more accurately for the unknown sets.

After the model has been trained by using the training set, it can be used for predicting the values of unknown features which are nothing but the testing set.

The testing set is passed for SVM prediction and the result of this set is the predicted values.

In this case it is the classified list of music instruments which are being played in the music track.

V. RESULTS AND DISCUSSION

The problem of classifying musical instruments is solved by this system. By using the SVM algorithm and performing operations on it will be able to generate classified musical instrument list. So, after the implementation of the project the output should come as successful classification of the instruments that are played in the audio file. The outcome should be with maximum accuracy in showing the instruments. Implementation on static dataset gave us the accuracy of 99% as shown in the Fig. 5. For the dynamic dataset 90% of accuracy is expected in the result.

```
In [19]: # Recall - the ability of the classifier to find all the positive samples
         print("Recall: ", recall_score(test_classes, predicted_labels,average=None))

         # Precision - The precision is intuitively the ability of the classifier not to
         #label as positive a sample that is negative
         print("Precision: ", precision_score(test_classes, predicted_labels,average=None))

         # F1-Score - The F1 score can be interpreted as a weighted average of the precision
         #and recall
         print("F1-Score: ", f1_score(test_classes, predicted_labels, average=None))

         # Accuracy - the number of correctly classified samples
         print("Accuracy: %.2f  ," % accuracy_score(test_classes, predicted_labels,normalize=True), accurac
         print("Number of samples:",test_classes.shape[0])

Recall:  [1.  1.  0.96 1.  1.  1. ]
Precision:  [1.         1.         1.         1.         0.96153846 1.        ]
F1-Score:  [1.         1.         0.97959184 1.         0.98039216 1.        ]
Accuracy: 0.99  , 149
Number of samples: 150
```

Fig. 5: Accuracy Result

The instruments which were used for the classification were cello, flute, oboe, sake, trumpet, viola. Confusion matrix is the raw result from automatic classification, presented in percentage. The values in diagonal cells are the accuracy of classification. For reading the value, system predicted 24 samples right out of 25 for Cello and it predicted 25 samples out of 25 right for Flute. Fig. 6 shows the Confusion Matrix for the results.

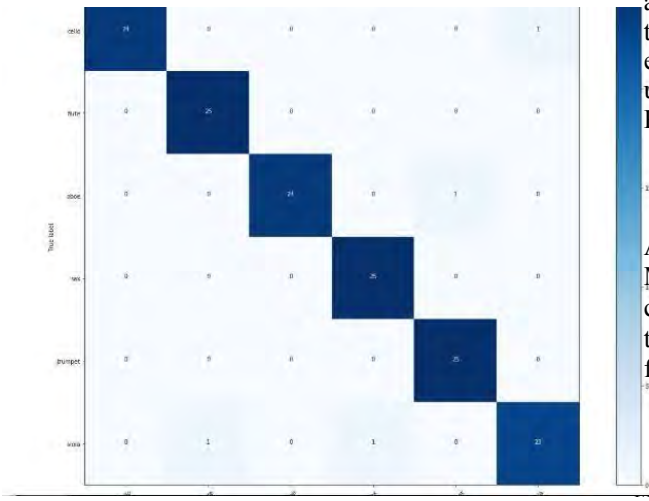


Fig. 6: Confusion Matrix

VI. Conclusion

When identifying musical instruments, selecting an indicative subset of features to predict on is extremely important. Selection of the most excellent contributing features to be extracted and the selection of the top suited method of classification are the most important decisions to be made for the content based audio classification. SVMs can be trained efficiently for audio classification. First, a group of training data is out there and can may be used to train a classifier. Second, once trained, the calculation in a SVM depends on a usually small number of supporting vectors and is speedy. Third, the distribution of audio data in the feature space is complex and different classes may have overlapping or interwoven areas. A kernel based SVM is well right to handle such a situation. SVM, implements mapping of inputs onto a high dimensional space with a group of non- linear basis functions. SVM can be used to study a variety of representations, such as neural nets, splines, polynomial estimators, etc, but there is a exclusive optimal solution for each selection of the SVM parameters. The results of the proposed scheme suggest the successful classification of musical instruments. Choosing the right feature for classifier can not only improve the accuracy, but also reduce attributes. Less attributes and higher accuracy are the basis of real-time automatic classification systems.

VII. Future Scope

The scope for exploring this system is endless. The future work can be done by combining MFCC with other features or using some feature selection algorithms to reduce attributes and eliminate similarity to get maximum marginal SVM. Less attributes and higher accuracy are the basis of real-time automatic classification systems.

Also Deep learning algorithms, Neural Networks algorithms when combined with feature extraction techniques gives the maximum accuracy. This can be explored for the real time systems and can also be used for different applications of Music Instrument Retrieval.

Acknowledgment

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra, the Vice-Principal, Dr. Deven Shah and the college management for their support. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

References

[1] ohjyot Singh, Dnyaneshwar Bachhav, Omkar Joshi, Nita Patil, usical Instrument Recognition using CNN and SVM", International Research Journal of Engineering and Technology (IRJET) Volume: 06 Issue: 03 | Mar 2019.

[2] Miss Karishma Vichare, Prof. S. R. Gulhane, "Musical Instrument Identification and status fonding with MFCC", International Journal of Advance Engineering and Research Development Volume 5, Issue 04, April -2018.

[3] Lewis Guignard and Greg Kehoe, "Learning Instrument Identification", Autumn 2015, Stanford University.

[4] Jing Liu and Lingyun Xie, "SVM-Based Automatic Classification of Musical Instruments", IEEE Int. Conf. June 2010.

[5] Taejin Park and Taejin Lee, "Musical Instrument Sound Classification with Deep Convolutional Neural Network using Feature Fusion Approach", Electronics and Telecommunications Research Institute (ETRI), Republic of Korea.

[6] Ms. Shabnam R. Makandar, “Music Classification using Support Vector Machine”, International Journal of advance Research in Science and Engineering Volume No. 07, Issue No. 3, March 2018.

[7] A. S. Patankar, “Musical Instrument Recognition Techniques”, International Journal Of Researches In Biosciences, Agriculture And Technology, Vol. V, Issue (1), Jan.- 2017.

[8] Janet Marques and Pedro J. Moreno, “A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines”, June 1999.

[9] Slim Essid, Gaël Richard, Member, IEEE, and Bertrand David, “Musical Instrument Recognition by Pairwise Classification Strategies”, IEEE Transactions On Audio, Speech, And Language Processing, VOL. 14, NO. 4, JULY 2006.

[10] Musical Instrument Sound Classification by R Rayar, M Anto Bennet, A Nazreen Banu, A Sushanthi, M Rajasekar.

[11] Elzbieta Kubera, Alicja A Wieczorkowska, and Magdalena Skrzypiec. Influence of feature sets on precision, recall, and accuracy of identification of musical instruments in audio recordings. In ISMIS, pages 204–213. Springer, 2014.

[12] Weston, J. and Watkins, C. (1998). "Multi-class Support Vector Machines," Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May.

[13] Wikipedia contributors, Musical Instrument Classification [OnLine]. Available:http://en.wikipedia.org/w/index.php?title=Musical_instrument_classification&oldid=325554329.

[14] Stan Z. Li Lie Lu and Hong-Jiang Zhang. Content-based audio segmentation using support vector machines.

[15] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313 :504–507, 2006.

[16] A. A. Livshin, and X. Rodet, Musical instrument identification in continuous recordings, in Proc. Int. Conf. Digital Audio Effects,, Italy, 2004.

[17] C.J.C.Burges,“Atutorialonsupportvectormachinesforpatternrecognition,” J. Data Mining Knowledge Discovery, vol. 2, no. 2, pp. 1–43, 1998.

[18] https:/analyticsvidhya.com/

[19] Akram Azarloo, Fardad Farokhi,”Automatic Musical Instrument Recognition Using K-NN and MLP Neural Networks”, Faculty of Electrical and Electronics Engineering Islamic Azad University,2012 Fourth International Conference on Computational Intelligence, Communication Systems and Networks.

[20] Swati D. Patil1 and Tareek M. Pattewar,”Musical Instrument Identification Using SVM and MLP with Formal Concept Analysis”.

[21] S. R. Sankaye and U. S. Tandon, “Human Perception and Recognition of Musical Instruments”.

[22] Jair Cervantes, Xiaoou Li and Wen Yu, “SVM Classification for Large Data Sets by Considering Models of Classes Distribution,”Proc. Sixth Mexican International Conference on Artificial Intelligence-Special Session (MICAI 2007), 2007, pp.51-60, doi: 10.1109/MICAI.2007.27.

[23] Youtube to mp3 generator. www.youtube-mp3.org.

[24] Y Takahashi and K Kondo. Comparison of two classification methods for musical instrument identification. In Consumer Electronics (GCCE), 2014 IEEE 3rd Global Conference on, pages 67–68. IEEE, 2014.

[25] C. Pruisers,I. Kaminskyj and Schnapp, Wavelet analysis in musical instrument sound classification, in international symposium on signal processing and its applications, pp. 14, 2005.

Design and Development of System for Removal of Background in Images using Semantic Segmentation

Saurabh. B . Jha,Saurabh. P. Jha,Meetu Bhardwaj ,Dr. Sheetal Rathi
Computer Engineering
TCET Mumbai, India
saurabh.jha009@gmail.com, saurabhjha57@gmail.com meetubhardwaj15@gmail.com ,

Abstract— **Image segmentation is that the task of partitioning a picture into multiple segments. This makes it a whole lot easier to analyze the given image. And essentially, that is what we are always striving for in computer vision. This is quite almost like grouping pixels together on the idea of specific characteristic(s). Now these characteristics can often lead to different types of image segmentation. Image segmentation may be a longstanding computer Vision problem. Quite a few algorithms have been designed to solve this task, such as the Watershed algorithm, Image thresholding, K-means clustering, Graph partitioning methods, etc. removal of background images also require image segmentation where we divide the fore image with background image and gets it available for analysis. Removing the background from the image makes the fore image easy to study and becomes much useful for the e-commerce sites for sale where the images are required to have white back**

Keywords—Artificial Intelligence, Segmentation, Image.

I. INTRODUCTION

Image segmentation is that the division of an image into regions or categories, which correspond to different objects or parts of objects. Every pixel during a picture is allocated to a minimum of one among sort of those categories. an honest segmentation is typically one during which, pixels within an equivalent category have similar greyscale of multivariate values and form a connected region and neighbouring pixels which are in several categories have dissimilar values. The results of image segmentation could also be a group of segments that collectively cover the entire image, or a gaggle of contours extracted from the image (see edge detection). Each of the pixels during a neighbourhood are similar with regard to some characteristic or computed property, like colour, intensity, or texture. Adjacent regions are significantly different with regard to the same characteristic(s). When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation are often used to create 3D reconstructions with the help of interpolation algorithms like marching cubes. Having no background within the image helps to know the foe image properly and helps the customer visiting the e-commerce sites view the merchandise more appropriately and evaluate it properly before making any purchase. E-commerce sites make sure the product available at their websites are authentic and of excellent quality. Customers don't have privilege to see the merchandise physically like old fashioned shopping in order that they need to make sure the product is clearly visible to them through pictures altogether different angles possible. During this any kind of background may create complexity in understanding the merchandise so it's best to get rid of the background from the image.

II. ALGORITHM AND DESIGN PHASE

Block Diagram

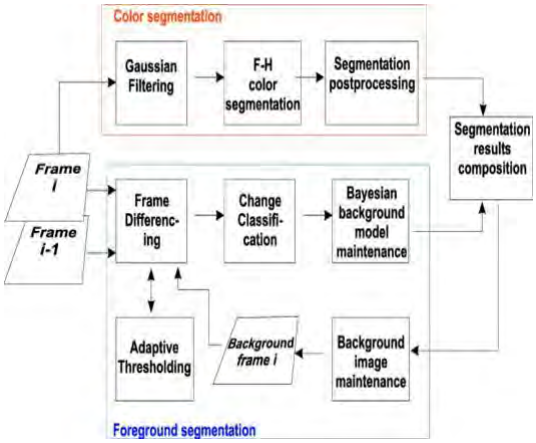


Fig: General System Block Diagram

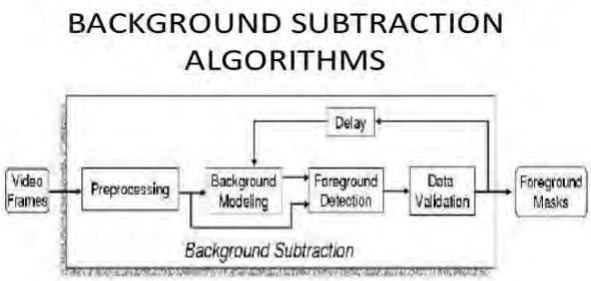
Description:

In the overall system the image get divided into multiple frames which later goes under differencing and classification for experience learning and takes a background maintenance to differentiate fore and background image. Further it goes under segmentation and is restored using thresholding

3
4

Concept.

Work Flow Diagram



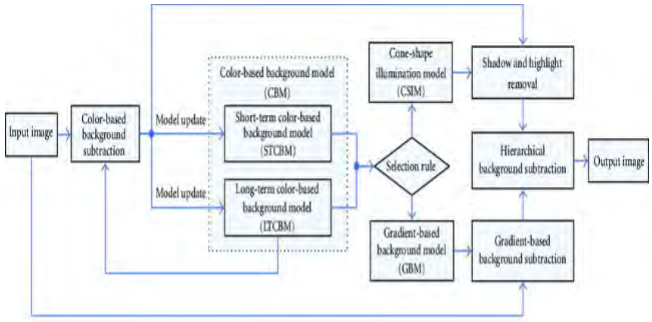
Flow diagram of a generic background subtraction algorithm.

Fig: Background Subtraction Work Flow

Description:

It collects all the frames, pre-processes it as per the learning algorithm. Models the entire background image, detects the fore image separates them visualizes them and provides the fore the fore mask overlay.

Algorithm/Flowchart



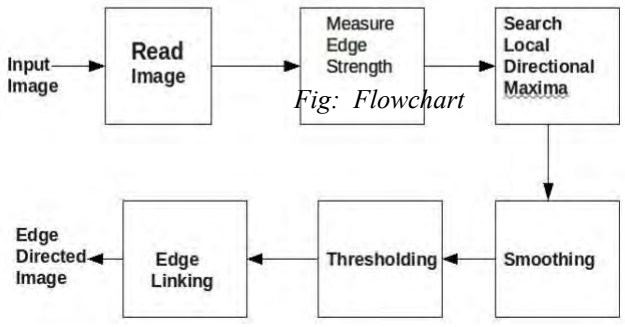


Fig: Flowchart

Fig: Edging Flowchart

III. PROPOSED WORK

Cleaning: -To remove noise, we use a simple median filter to remove the outliers, but one can use a different noise removal approach or artefact removal approach. The artefact vary across acquisition systems (microscopy techniques) and may require complicated algorithms to restore the missing data. Artefacts commonly fall into two categories:

- blurry or out-of-focus areas
- imbalanced foreground and background (correct with histogram modification)

Segmentation: -For this article, we limit segmentation to Otsu's approach, after smoothing an image using a median filter, followed by validation of results. You can use an equivalent validation approach for any segmentation algorithm, as long because the segmentation result's binary. These algorithms include, but aren't limited to, various Circular Thresholding approaches that consider different color space.

Some examples are:

- Li Thresholding
- An adaptive thresholding method that is dep indent on local intensity
- Deep learning algorithms like UNet used commonly in biomedical image segmentation
- Deep learning approaches that semantically segment an image

Validation: -We begin with a ground truth data set, which has already been manually segmented. To quantify the performance of a segmentation algorithm, we compare ground truth with the anticipated binary segmentation, showing accuracy alongside simpler metrics. Accuracy are often abnormally high despite a coffee number of true positives (TP) or false negatives (FN). In such cases, F1 Score and MCC are better quantification metrics for the binary classification. We'll go into detail on the pros and cons of these metrics later. For qualitative validation, we overlay the confusion matrix results i. e where exactly the true positives, true negatives, false positives, false negatives pixels are onto the grayscale image. This validation also can be applied to a color image on a binary image segmentation result, although the info we utilized in this text may be a grayscale image. In the end, we'll present the entire process in order that you'll see the results for yourself. Now, let's look at the data –and the tools used to process that data.

Loading and visualizing data

We will use the below modules to load, visualize, and transform the data. These are useful for image processing and computer vision algorithms, with simple and complex array mathematics. The module names in parentheses will help if installing individually.

FIG (importing modules)

Individual tiles can be mapped to run on
Multiprocessing/multithreaded (i.e. distributed
Infrastructure), then stitched back together to get the
Complete segmented image.
Fig loading and visualizing

Pre-processing
Before segmenting the info, you ought to undergo the dataset thoroughly to work out if there are any artifacts thanks to the imaging system. In this example, we only have one image in question. By watching the image, we will see that there aren't any noticeable artifacts that might interfere with the segmentation. However, you'll remove outlier noise and smooth a picture employing a median filter. A median filter replaces the outliers with the median (within a kernel of a givensize).

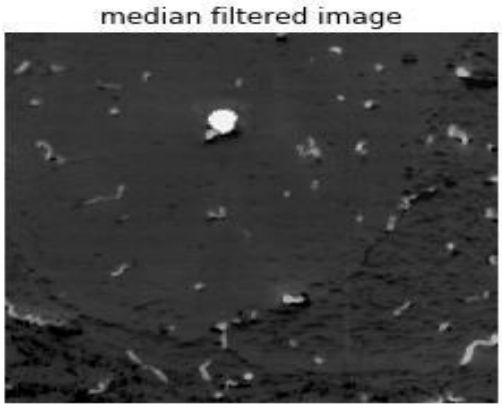


Fig: Median Filter Image

To determine which thresholding technique is best for segmentation, you could start by thresholding to determine if there is a distinct pixel intensity that separates the two classes. In such cases, you can use that intensity obtained by the visual inspection to binarize the image. In our case, there seem to be a lot of pixels with intensities of less than 50 which correspond to the background class in the inverted grayscale image.

Although the distribution of the classes isn't bimodal (having two distinct peaks), it still features a distinction between foreground and background, which is where the lower intensity pixels peak and then hit a valley. This exact value areoften obtained by various thresholding techniques. The segmentation section examines one such method intimately. *Visualize histogram of the pixel intensities*

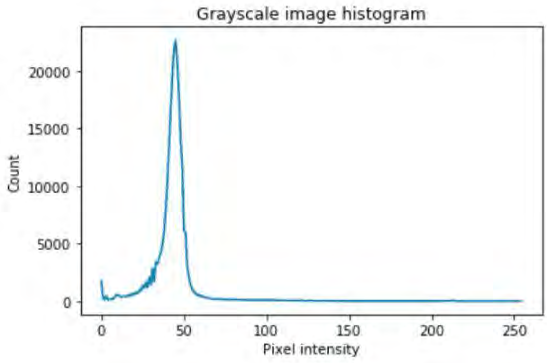


Fig: Grayscale Image Histogram

IV. REQUIREMENTS

Software Requirements-
- Front End-HTML, CSS, JavaScript o Back Python, Jupyter, Spyder
- Operating System

Hardware Requirements-
- Min Memory: 4GB o Disk size: 1TB
- Processor: 64bit CPU with 8 cores and speed of 2.4 GHz

User Requirement –

- Interface for inputting image o Output alteration space
- Output collection space

V. METHODOLOGY

The purpose of developing this project is to upgrade the automatic removal technique and make the process much faster. In this AI helps the user to make the output as per their choice and alter the specification as per their choice. User enabled camera or platform to run the system is to be compatible to receive desired output. Since the system is getting developed by artificial intelligence system keeps learning from the previous images and learning given.

It enables all the libraries required for the processing. Every iteration it visualizes the data to ensure the smooth procedure flow.

Segmentation: -
After removing noise, you'll apply the sk-image filters module to undertake all thresholds to explore which thresholding methods fare well. Sometimes, in a picture, a histogram of its pixel intensities isn't bimodal. So, there could be another thresholding method which will fare better like an adaptive thresholding method that does thresholding supported local pixel intensities within a kernel shape. It's good to ascertain what the various thresholding methods

results are, and sk-image filters thresholding try all threshold () is handy for that.

The simplest thresholding approach uses a manually set threshold for a picture. On the opposite hand, using an automatic threshold method on a picture calculates its numerical value better than the human eye and should be easily replicated. For our image during this example, it looks like Otsu, Yen, and therefore the Triangle method are performing well. the opposite results for this case are noticeably worse.

We'll use the Otsu thresholding to segment our image into a binary image for this text. Otsu calculates thresholds by calculating a worth that maximizes inter-class variance (variance between foreground and background) and minimizes intra-class variance (variance within foreground or variance within background). It does well if there's either a bimodal histogram (with two distinct peaks) or a threshold value that separates classes better.

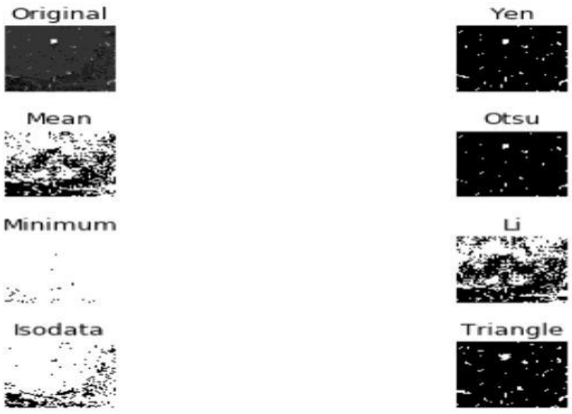


Fig: Otsu Visualization

Otsu thresholding and visualization

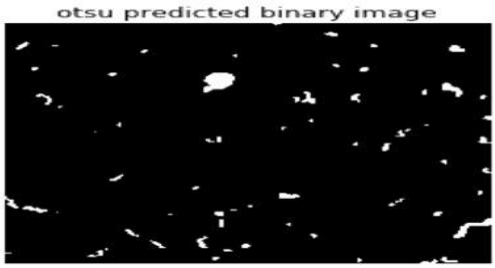


Fig: Otsu Binary Image

If the above simple techniques don't serve the aim for binary segmentation of the image, then one can use UNet, ResNet with FCN or various other supervised deep learning techniques to segment the images. To remove small objects thanks to the segmented foreground noise, you'll also consider trying skimage.morphology.remove_objects(). Validation

In any of the cases, we'd like the bottom truth to be manually generated by a person's expertly within the image type to validate the accuracy and other metrics to see how well the image is segmented.

The confusion matrix

We use sklearn.metrics.confusion_matrix() to urge the confusion matrix elements as shown below. Scikit-learn confusion matrix function returns 4 elements of the confusion matrix, as long as the input may be a list of elements with binary elements. For edge cases where everything is one binary value(0) or other(1), sklearn returns just one element. We wrap the sklearn confusion matrix function and write our own with these edge cases covered as below:

Groundtruth

Predicted	Foreground (255, vessels)	Background (0, tissue)
Foreground (255,vessels)	True Positive (TP)	FP (False Positive)
Background (0, tissue)	False Negative (FN)	TN (True Negative)

Fig: Image Confusion Matrix

Accuracy: - -Accuracy is a common validation metric in case of binary classification. It is calculated as

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Fig:Accuracy Formula

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

It varies between 0 to 1, with 0 being the worst and 1 being the best. If an algorithm detects everything as either entirely background or foreground, there would still be a high accuracy. Hence, we'd like a metric that considers the imbalance in school count. Especially since the present image has more foreground pixels (class 1) than background 0.

F1 score

The F1 score varies from 0 to 1 and is calculated as:

$$\frac{2*TP}{2*TP + FP + FN}$$

Fig: *F1 Score value*

with 0 being the worst and 1 being the simplest prediction. Now let's handle F1 score calculation considering edge cases. An F1 score of above 0.8 is considered a good F1 score indicating prediction is doing well.

MCC**:** -*MCC* stands for Matthews Correlation Coefficient, and is calculated as:

$$\frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Fig: Mathhews Correlation Coefficient

It lies between -1 and +1. -1 is absolutely an opposite correlation between ground truth and predicted, 0 is a random result where some predictions match and +1 is where absolutely everything matches between ground and prediction resulting in positive correlation. Hence, we need better validation metrics like MCC.

In MCC calculation, the numerator consists of just the four inner cells (cross product of the elements) while the denominator consists of the four outer cells (dot product of the) of the confusion matrix. In the case where the denominator is 0, MCC would then be ready to notice that your classifier goes within the wrong direction, and it might notify you by setting it to the undefined value (i.e. numpy, nan). But, for the aim of getting valid values, and having the ability to average the MCC over different images if necessary, we set the MCC to -1, the worst possible value within the range. Other edge cases include all elements correctly detected as foreground and background with MCC and F1 score set to 1. Otherwise, MCC is about to -1 and F1 score is 0.

To learn more about MCC and therefore the edge cases, this is often an honest article. To understand why MCC is better than accuracy or F1 score more in detail, Wikipedia does good work here.

Finally, we will compare the validation metrics by result, side-by-side.

Accuracy is close to 1, as we have a lot of background pixels in our example image that are correctly detected as background (i.e. true negatives are are naturally higher). This shows why accuracy isn't an honest measure for binary classification.

F1 score is 0.84. So, during this case, we probably don't need a more sophisticated thresholding algorithm for binary segmentation. If all the pictures within the stack had similar histogram distribution and noise, then we could use Otsu and have satisfactory prediction results.

The MCC of 0.85 is high, also indicating the ground truth and predicted image have a high correlation, clearly seen from the predicted image picture from the previous section. Now, let's visualize and see where the confusion matrix elements TP, FP, FN, TN are distributed along the image. It shows us where the threshold is picking up foreground (vessels) when they are not present (FP) and where true vessels are not detected (FN), and vice-versa.

Validation visualization

To visualize confusion matrix elements, we find out exactly where within the image the confusion matrix elements fall. For example, we find the TP array (i.e. pixels correctly detected as foreground) is by finding the logical "and" of the bottom truth and therefore the predicted array. Similarly, we use logical boolean operations commonly called as Bit blit to find the FP, FN, TN arrays.

Then, we can map pixels in each of these arrays to different colors. For the figure below we mapped TP, FP, FN, TN to

The CMYK (Cyan, Magenta, Yellow, Black) space. One could similarly also map them to (Green, Red, Red, Green) colors. We would then get a picture where everything in red signifies the wrong predictions. The CMYK space allows us to differentiate between TP, TN.

We use opencv here to overlay this color mask onto the first (non-inverted) grayscale image as a transparent layer. This is Called Alpha compositing:

Fig: Overlay Mask

VI. TECHNOLOGY USED

1) Python

Python is an interpreter, object-oriented, high-level programing language with dynamic semantics. Its high-level inbuilt data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, also as to be used as a scripting or glue language to connect existing components together. Python's simple, easy to find out syntax emphasizes readability and thus reduces the value of program maintenance. Python supports modules and packages, which inspires program modularity and code reuse. The Python interpreter and therefore the extensive standard library are available in source or binary form for free of charge for all major platforms, and may be freely distributed.

2) *Jupyter Notebook*

The Jupyter Notebook is a a living online notebook, letting faculty and students weave together computational information (code, data, statistics) with narrative, multimedia, and graphs. Faculty can use it to line up interactive textbooks, filled with explanations and examples which students can test out right from their browsers. Students can use it to elucidate their reasoning, show their work, and draw connections between their classwork and therefore the world outside. Scientists, journalists, and researchers can use it to open up their data, share the stories behind their computations, and enable future collaboration and innovation

3) *Spyder:*

A spider is a software program that travels the online (hence the name "spider"), locating and indexing websites for search engines. All the main search engines, like Google and Yahoo!, use spiders to create and update their indexes. These

Programs constantly browse the online, traveling from one hyperlink to a different.

VII. CONCLUSION

The current system being used is developed using advanced photoshop due to which accuracy is not obtained whereas the new system is developed using Artificial Intelligence helping to attain optimization. Segmentation will make the task much faster than the manual technology. Process of bulk images processing can be thought due to this. It will help the seller and buyer both on the e-commerce webpages as the seller can easily and quickly obtain the desired image. Buyer can evaluate the product more accurately due the precise image uploaded. Artificial Intelligence makes the process faster and quicker just as required in today's world

References

[1] Linda G. Shapiro and George C. Stockman (2001): "Computer Vision", pp 279-325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3..

[2] Barghout, Lauren, and Lawrence W. Lee. "Perceptual information processing system." Paravue Inc. U.S. Patent Application 10/618,543, filed July 11, 2003

[3] Batenburg, K J.; Sijbers, J. (2009). "Adaptive thresholding of tomograms by projection distance minimization". *Pattern Recognition*.

[4] Kashanipour, A.; Milani, N; Kashanipour, A.; Eghrary, H. (May 2008)

[5] Barghout, Lauren (2014). Vision. Global Conceptual Context Changes Local Contrast Processing (Ph.D. Dissertation 2003). Updated to include Computer Vision Techniques. Scholars' Press

[6] Background Removal in image indexing retrieval, Yi Lu, Hung Guo(IEEE)

[7] www.news.developer.nvdia.com/imageremoval

[8] www.semanticscholar.org/semanticsegmentation

[9] www.cloudinary.com/backgroundremoval

[10] https://shodhganga.inflibnet.ac.in/bitstream/10603/25143/9/09_chapter%2 03.pdf

Implementation of Intelligence Voice Analysis System Based On Acoustics and Speaking Fundamentals

Shailesh Sangle, Nirja Rajeev, Ankita Patel, Aditi Ramgaehia
Dept. of Computer Engineering
Thakur college of Engineering & Technology
Mumbai,India
shailesh.sangle@thakureducation.org, nirja.rajeev@gmail.com, ankspatelagp@gmail.com, ramgarhiaaditi@gmail.com

Abstract—**Voice is the basic and most used way of communication used for interaction. Numerous technologies enable machines to respond correctly and reliably to human voices and provide convenient and valuable services. Speech recognition is an useful application of Digital Signal Processing and has many real-world applications. Automatic Speech Recognition of gender, age group, emotion and disease from speech segments can be done using digital speech processing and pattern recognition techniques. Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. The main aim of this study is to create an accurate system for voice analysis. Acoustic coefficients were used to form test and reference templates.**

Keywords—voice analysis, voice feature extraction, speech processing, emotion recognition

I. INTRODUCTION

Interactive communication is an interface which involves the exchange of reciprocal ideas and emotions. Gestures and sound are a way of conveying information in a human-to- human interaction. Speech, a special form of sound, is one of the vital ways of conveying information between people. Words are not sole component of speech. Acoustic properties of speech also carry important affecting features. Emotions exist in every part of the speech. Emotions in speech are transmitted from one communicator to another during an interaction. As a result of exchange of emotions during an ongoing conversation, emotional state of a speaker may straightforwardly trigger an panelist emotional state resulting in a change in the speech style or tone.

Timbre, speech rate, and length are certain characteristics that discriminate human voices, mainly male and female voices. Distinction in gender can not be predicted by vocal speech. Some voice pitch may vary between male and female so it is difficult to predict male and female accurately. We can identify the gender of the respective speaker with the help of techniques used for speech processing in real time environment. Voiced fold thickness is the main reason behind the difference that can be calculated between the genders. Another reason that contributes to is the style in which the person speaks and the present physical situations.

II. FEATURES SELECTION

Many different speech feature extraction methods have been proposed over the years. Methods are distinguished by the ability to use information about human auditory processing and perception, by the robustness to distortions, and by the length of the observation window. Due to the physiology of the human vocal tract, human speech is highly redundant and has several speaker-dependent features, such as pitch, speaking rate and accent. An important issue in the design of a speech emotion recognition system is the extraction of suit-able features that efficiently characterize different emotions. Although there are many interesting works about automatic speech emotion detection there is not a silver bullet feature for this aim. Since speech signal is not stationary, it is very common to divide the signal in short segments called frames, within which speech signal can be considered as stationary. Human voice can be considered as a stationary process for intervals of 20 ms to 40 ms. If a feature is computed at each frame is called local, otherwise, if it is calculated on the entire speech is named global. There is not agreement in the scientific community on which between local and global features are more suitable for speech emotion recognition.*A.*

A.GENDER RECOGNITION FEATURES

Together with the Mel Frequency Cepstral Coefficients (MFCC) pitch is the most frequently used feature since it is a physiologically distinctive trait of a speaker's gender. Other employed features are formant frequencies and bandwidths, open quotient and source spectral tilt correlates energy between adjacent formants, fractal dimension and fractal dimension complexity jitter and shimmer (pitch and amplitude micro-variations, respectively) harmonics-to- noise-ratio, distance between signal spectrum and formants.

B. EMOTION RECOGNITION FEATURES

Coherently with the wide literature in the field, a set of 182 features have been analysed for each the recorded speech signal, including: Mean, adjustment, median, minimum, maximum and range of the amplitude of the speech Mean, variance, minimum, maximum and range of the formants; Energy of the Bark sub-bands Mean, variance, minimum, maximum and range of the Mel-Frequency Cepstrum Coefficients Spectrum shape features Centre of Gravity, Standard Deviation, Skewness and Kurtosis; Mean and standard deviation of the glottal pulse period, jitter local absolute, relative usual perturbation, difference of difference period and (–ve) point period perturbation quotient. various classification techniques like vector quantisation, dynamic time wrapping, support vector machine, Gaussian mixture model and artificial neural network. This paper described all these techniques in detail.

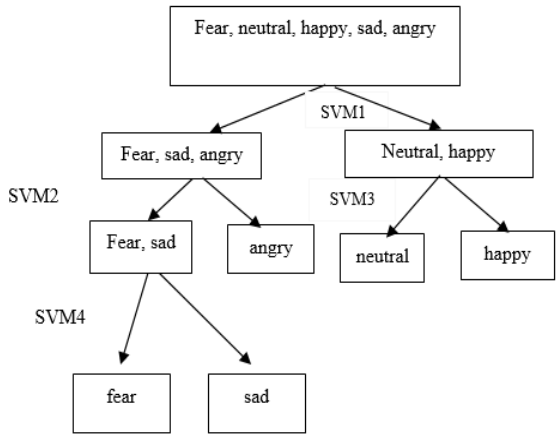


Figure 1: Overview of emotion analysis

C. DISEASE RECOGNITION

Voice features are extracted using some DSP techniques. These features contain information on health of voice tract and of the organs cooperating in speech production. These features represent the particular voice and may be used to discriminate voice of healthy and unhealthy persons. The time field analysis, spectrum analysis, cepstrum analysis, glottal waveform analysis are used to extract the voice

D.AGE GROUP RECOGNITION

Voice is a feature which changes with gender, age and state of mind. Features drastically change as we age. Higher pitch voice in men , lower pitch voice in women, reduced volume and projection of the voice(or thin voice) ,reduced vocal resolution , struggle being heard in noisy situations, tremor or shakiness in the voice, etc .These symptoms are amplified by the reduced hearing ability that commonly occurs in our peers as we age.

III. SYSTEM ARCHITECTURE

We have two phases: testing and training phases. We sampled and tested hundreds of voice samples in the testing phase. All the results of these voice samples are stored in the database. The human voice goes through the following steps: Features of voice are extracted these features are passed through different algorithms to find out the different results we want. First we pass the features through gender identification model, then through age classification algorithm and then emotion. From the testing phase, a feature model is created. This feature model is used for prediction of voices in the training phase.features. These features are then classified into groups using

IV.TECHNOLOGY USED

[1]MFCC

In sound processing, the mel-frequency cepstrum is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log

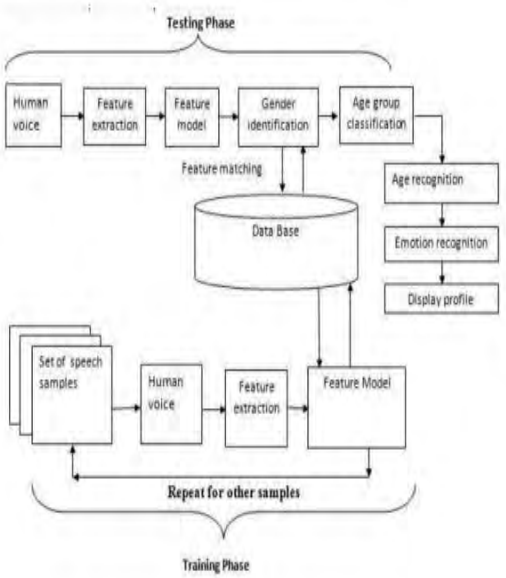


Figure 2: System architecture

power spectrum on a nonlinear mel scale of frequency. Mel- frequency cepstral coefficients are coefficients that collectively make up an MFC. It is used to extract different features of voice like pitch, frequency ,loudness,etc.

[2] HMM

Modern general-purpose speech recognition systems are based on Hidden Markov Models.HMMs are used in speech recognition because a speech signal can be viewed as a short-time stationary signal. In a short time- scale, speech can be approximated as a stationary process.Another reason why HMMs are popular is because they can be trained repeatedly and are simple and computationally feasible to use.We have used HMM in gender identification. [3] GMM Gaussian Mixture Modelling are basically used to model the speech feature distribution.We have used it along with HMM.

[4] Tkinter

Python provides many methods of creating a Graphical User Interface GUI. Tkinter is one of the easiest and most popular methods of creating a GUI in python.It is an in- built Tk GUI toolkit provided in python.Due to easy programming we have used tkinter to develop the GUI.

[5] Pygame

Pygame is a cross-platform set of Python modules mainly designed for writing video games.It also has sound libraries designed to be used with the Python programming language.

[6] SciPy

SciPy is an open-source Python library used for scientific computing and technical computing. SciPy contains modules for FFT, signal and image processing,We have used it along with MFCC to extract voice features.

[7] Librosa

Librosa is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems

V. FLOWCHART

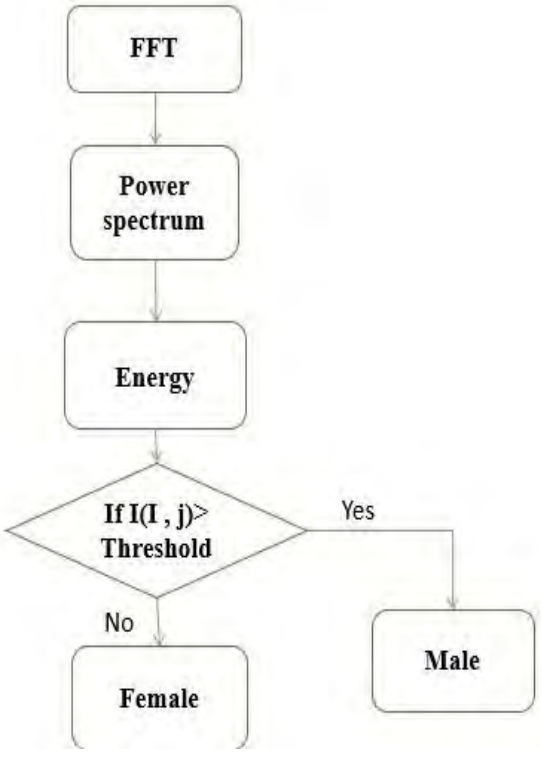


Figure 3:Gender identification

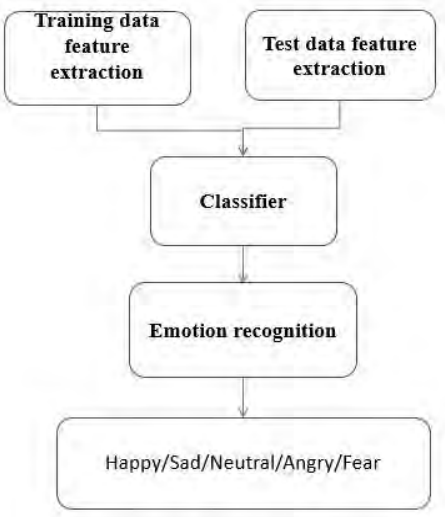


Figure 4:Emotion identification

VI. RESULTS

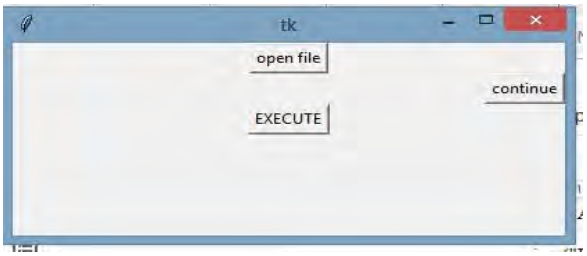


Figure 5:GUI

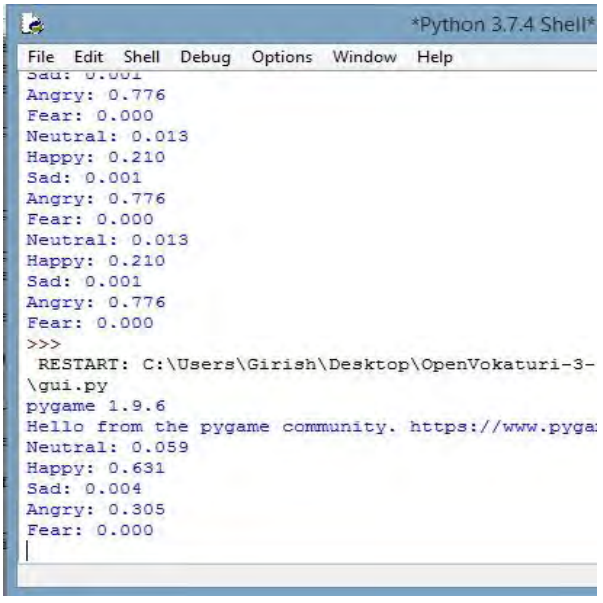


Figure 6:Identified emotions

VII. CONCLUSION

This review paper is to create a smart system which can be used in client interaction ,law implementation,speaker identification and psychological analysis. This can be achieved by a easy microphone transducer with some programming

involved. Thousands of inputs can be given and the voice acknowledgement can be achieved to find the age range of a person who has spoken. This can also provide cyber security. Further modifications can be made for the same. The proposed system, able to recognize the emotional state of a person starting from acoustic signals registrations,is composed of 4 functional blocks: Gender Recognition (GR) ,age group recognition(AGR),Disease recognition(DR), Emotion Recognition (ER). They have been implemented by a Pitch Frequency Estimation method, Support Vector Machine ,MFCC,ANN etc.(SVM) classifiers (fed by properly selected audio features), which exploit the GR subsystem output. The performance analysis shows the accuracy obtained with the adopted emotion recognition system in terms of recognition rate and the percentage of correctly recognized emotional contents. The experimental results highpoint that the Gender Recognition (GR) subsystem allows increasing the overall emotion recognition .This happens due to the prior knowledge of the speaker gender.The results show that with the employment of a features selection algorithm, a satisfying recognition rate level can still be obtained also reducing the employed features and, as a con-sequence, the number of operations required to identify the emotional contents. This makes feasible future development of the proposed solution over mobile devices.

VIII. ACKNOWLEDGMENT

We would like to thank all our professors who helped us in making this review paper successful. We thank Mr. Shailesh Sangle for guiding us in progressing in our paper. We thank TCET for providing an opportunity for us to present our paper. We at last thank the almighty and our parents who helped us in achieving it.

References

[1] Dixit , Vikas Mittal , Yuvraj Sharma , Voice parameter analysis for disease detection ,IOSR Journal of Electronics and communication engineering (IOSR -JECE) e-ISSN :22782834 VOL 9 Issue 3 version I

[2] Poorna Banerjee Dasgupta ,Detection and Analysis of Human Emotions through Voice and Speech Pattern Processing International Journal of Computer Trends and Technology (IJCTT) – Volume 52 Number 1 October 2017

[3] Ms.Deepali S. Shrivastava(P.G Student), Prof. Saurabh Shah, Prof. Ketan Bhavsar ,Voice analysis based diseases identification using speech processing , January 2016 Volume 3 Issue 1. JETIR (ISSN-2349-5162)

[4] Yashpalsing Chavhan , M.L Dhore,Pallavi Yesaware, Speech emotion recognition using support vector machine, 2010 International Journal ofComputer Applications 0975- 8887 (Volume 1 No. 20)

[5] KC Toran , Lal BK, Objective analysis of voice in normal young adults , Kathmandu University Medical Jornal(2009), Vol. 7 , No 4, Issue 28, 374-377

[6] A Raahul,R Sapthagiri, K Pankaj and V Vijayarajan , Voice based gender classification using machine learning ,IOP Conf. Series: Materials Science and engineering 263 (2017)

[7] Gil Dobry “Speaker age estimation based on acoustic speech signal” The open University of Israel Department of Mathematics and Computer Sciences

Implementing System Forsign Languagerecognitionusing Machinelearning

Abhijeet Singh,Ankit Singh,Rashmi Singh,Nitin Harane
ComputerEngineering,TCET
University of Mumbai
ohhyessabhii99@gmail.com,ankii4050@gmail.com,ankii4050@gmail.com,nitinharane@thakureducation.org

***Abstract*—Handgesturesareusedasaway forpeopleto expressthoughtsandfeelingsit servestoreinforce informationdeliveredinourdaily conversation.Sign languageisa structuredformofhandgesturesinvolving visual motionsand signs, whichareusedasacommunication system. Forthedeafandspeech- impaired community,signlanguage servesasusefultoolsfordaily interaction.Signlanguageinvolvestheuseofdifferentparts ofbodynamelyfingers, hand,arm, head,bodyandfacial expressiontodeliverinformation.Thisposesa genuine communication barrierbetweenthedeafcommunityand therest ofthesociety,asaproblemyettobefullysolved untilthisday.**

***Keywords*— CNN(ConvolutionalNeuralNetwork),Hand motioncomparison,Signlanguage,Communication**.

I.INTRODUCTION

Handgestureisoneofthemethodusedinsignlanguagefor non-verbal communication.Itismostcommonlyused by deaf anddumb peoplewho have hearing orspeechproblemsto communicate among themselves or with normal people. Varioussign languagesystemshasbeendevelopedbymany makersaroundtheworldbuttheyareneitherflexiblenorcost. Effectiveforthe endusers.Patternrecognitionandgesture recognitionarethe developing fieldsofresearch.Being a significantpartinnonverbal communicationhandgesturesare playing key role in ourdaily life.Hand gesturerecognition systemprovides us aninnovative,Natural,Userfriendlywayof communicationwiththecomputerwhichismorefamiliartothe humanbeings.Byconsideringinmindthesimilarities ofhuman hand shapewithfourfingersandonethumb,Thesoftwareaims topresentarealtimesystemforrecognitionof handgestureon basisof detectionof some shapebasedfeatureslikeorientation, Centreofmass centroid,Fingerstatus,Thumb in position of raisedorfoldedfingers ofhand.

II.HOWITWORKS

Notebookdocuments(or “notebooks”,alllowercase)are documentsproduced bytheJupiterNotebook App,which containbothcomputercode(e.g.python)andrichtextelem ents (paragraph,equations,figures,links.Notebookdocument s are bothhuman-readabledocumentscontaining theanalysis descriptionandtheresults(figures,tablesaswell asexecutable documentswhichcan beruntoperformdataanalysis.The JupiterNotebook Appisaserver-clientapplicationthatallows editingandrunningnotebookdocumentsviaa w e b b r o w s e r

TheJupiterNotebookAppcan beexecutedonalocal desktop requiring nointernetaccess(asdescribedinthisdocument)or canbeinstalledonaremote server andaccessedthroughthe internet.TheJupiterNotebook isanopen-sourceweb applicationthatallowsyoutocreate andsharedocumentsthat containlivecode, equations, visualizationsandnarrativetext. Usesinclude:datacleaningandtransformation, numerical simulation, statisticalmodeling, datavisualization, machine learning, and much more. JupiterNotebook is a web applicationthat allows youtocreateandsharedocuments.

A. CNN

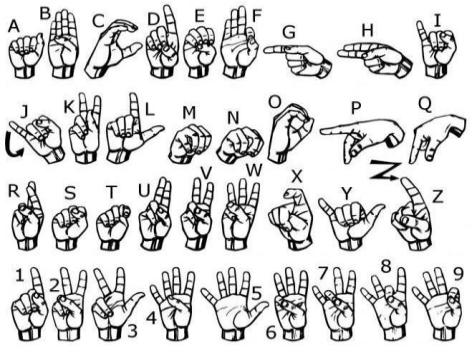
SignLanguageRecognitionusesneuralnetwork called3D ConvolutionalNeuralNetwork(CNN).The3DCNNmethodis based on2DCNNwhichisusedinspeechandimage recognition.The2DCNNimplementationis donebyextracting alayerto featuremap byusingkernelwindowcalled localreceptive field.Itreducesthe freevariablesandincrease generalizationcapabilityofthe network.The3DCNNis implementedusing2DCNNbyaddingthe motionfeature.The3D Convolutionisachievedbyconvolving3Dkernelwiththe cubeformedbystackingmultipleneighboringframestogether.12uses two3D-CNNintheirresearch,forextractinghandand

Upperbodyfeatures. EachCNNis3layersdeep.Theyalso applied localcontrastnormalization (LCN) forthefirsttwolayerandtheneuronsarerectifiedlinearunits(ReLU).2 1uses

2DConvolutional NeuralNetworkintheirresearch,with unusual featureintheirnetworkarchitecture.Theunusual features are ReLU Nonlinearity, multiple GPUs, LocalResponseNormalization, andOverlappingPooling.TheReLU Nonlinearityisthealternativeoftraditional neuronmodelsin CNN.Ittrainsseveral timesfasterthantraditional CNNandis bestfittobeusedinabigdataset.TheyalsousemultipleGPUs totraintheirdatabecauseofthesizeofdatathatdonotfitin oneGPU.

B. Dataset

Thedatasetformatispatternedtomatchcloselywiththeclassic MNIST.Eachtrainingandtestcaserepresentsalabel (0–25)as aone-to-onemapforeachalphabeticletterA-Z (andnocases for9=Jor25=Zbecauseofgesturemotions).Thetraining data (27,455cases)andtestdata(7172cases)areapproximatelyhalf the sizeof the standardMNISTbut otherwisesimilarwitha headerrow ofthelabel,pixel1,pixel2….pixel784 which representasingle28x28 pixelimagewithgrayscalevalues between0–255.



C. DataPreprocessing

Asthe datasethasalreadygivenCSVvaluesforimages,we don’tneedtodomuchpreprocessing.Ifdatasetoftheimage wasinrawformat, we have toconvertheminCSVformat arraysbefore doing anyofthefurtheroperations.Still,we performthefollowingsteps:

•Separatefeatures (784pixelcolumns)andoutput (resultlabel)
•Reshapethefeatures

D. Model

WewilluseKerastobuildthesimpleCNN(Convolutional NeuralNetwork).
Therearetotal7layersintheCNN:

1.1stConvolutional Layerwithrelu
2.1stMaxPooling
3.2nd ConvolutionalLayerwithrelu
4.2ndMaxPooling
5. Flattening
6. FirstFullLayerwithrely
7.OutputLayerwithsigmoid

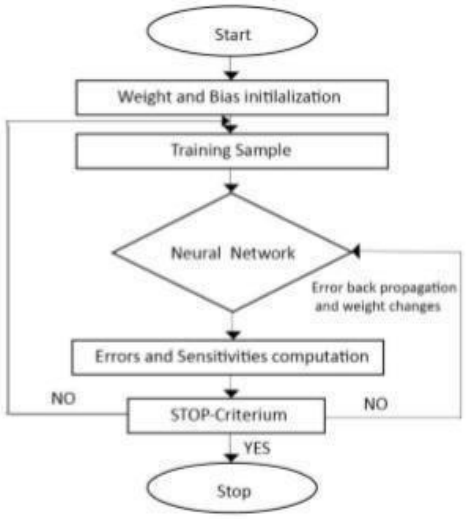
E. OpenCV

CreateaWindow:-
Wehavetocreateawindowtotakethe inputfromourwebcam.Theimagewhichwearetakin g asan inputshouldbe28x28 grayscaleimage.Becausewetrainedour modelon28x28 sizeimage.

F. Applications

Infuturework,proposedsystemcan be developedand implementedusing RaspberryPi.ImageProcessing partshould be improvedsothatSystemwouldbe able tocommunicate in bothdirectionsi.e.itshouldbecapable ofconverting normal languagetosignlanguageandviceversa.Wewill tryto recognize signswhichincludemotion. Moreover wewillfocus onconverting thesequence ofgesturesintotexti.e.wordand sentencesandthenconvertingitintothe speechwhichcanbe heard.

I. Flowchart



FigureI:-Implementationsteps carriedinthesystem

CONCLUSION

Signlanguagegesturerecognitionasit will behelpful todeaf people inassistingthemtointeractwithcomputers.However,it wasnoteasyformostoftheparticipantstoplacetheirhandon

Therectangular box,whichwasconsideredtobetheregionof recognition. Henceitisappropriatetorecommendthatabigger areaisrequiredsothattheparticipantswillnothavetostruggle toplacethe handintheregionbeforetheystartperforming the gestures.Forthesystemtobeeffectivelyappliedinacontrolled environmentlikeadeaf classroomthenawhitebackgroundcan beplacedbehindeveryseat of thecomputer user toreducethe challengeof complexbackgroundnoiseandimprovethe backgroundsubtraction.

ACKNOWLEDGMENT

Wesincerelythankour guideMr.NitinHaraneforhis guidance andconstantsupportandalsoforthesticktoourbacks.We alsothank the projectcoordinatorsforarrangingthe necessary facilities tocarryouttheprojectwork.

References

[1]Panwar.M. HandGesture Recognition System based on Shapeparameters.InProc.InternationalConference, Feb2012.
[2]Christopher Leeand YangshengXu. Online, interactivelearningofgestures forhumanrobotinterfaces. CarnegieMellonUniversity,TheRoboticsInstitute,Pittsburgh, Pennsylvania,USA,1996.
[3]Hyeon-KyuLeeandJinH.Kim.AnHMM-BasedThreshold ModelApproachforGestureRecognition.IEEE transactionsonpattern analysis and machine intelligence, Volume21,October1999.
[4]P.SubhaRajamandDr.G. Balakrishnan.RealTimeIndian SignLanguage RecognitionSystemtoaidDeafdumbPeople. ICCT,IEEE,2011.
[5]OlenaLomakina. Development of Effective Gesture RecognitionSystem.TCSET'2012,Lviv-Slavske,Ukraine, February21-24,2012.
[6]EtsukoUeda,YoshioMatsumoto,MasakazuImai,Tsulasa Ogasawara.HandPoseEstimationforVisionbased HumanInterface.InProc.10thIEEEInternationalWorkshop onRobot andHumanCommunication(Roman 2001),pp.473-478,2001
[7]Claudia Nölkerand HelgeRitter. Visual Recognitionof HandPostures.InProc.InternationalGesture

Workshop on Gesture-Based Communication inHuman

ComputerInteraction,pp. 61-72,1999.

[8]MeenakshiPanwarandPawanSinghMehra.HandGestu re

RecognitionforHuman ComputerInteraction.In

Proc. IEEE InternationalConference on Image Information

Processing,Waknaghat,India,November2011.

[9]Sanjay Meena. A Study ofHand Gesture Recognition

Technique.MasterThesis,DepartmentofElectronics and CommunicationEnginnering,NationalInstituteofTechno logy, India,2011.

Enhanced Text Summarization System for Key Data Extraction from Research Papers

Shrutika Agrawal, Saksham Khetan, Rrutum Lavana
Computer Department
Thakur College of Engineering and Technology
Mumbai, India
shrutikaagrawal188@yahoo.in, saksham.khetan@gmail.com, rrutum.lavana@gmail.com

Abstract— **in today's fast paced world, information on the internet is increasing at a very rapid rate. Because of this increasing information in the internet, text summarization is gaining more and more attention among the researchers. A system for text summarization is presented in this chapter. The specific problem under consideration is text summarization for key data extraction from research papers. This system aims to alleviate the burden of going through multiple lengthy research papers from the shoulders of the researchers by creating a summary of the desired research papers. The proposed system will make use of futuristic technologies like Natural Language Processing (NLP), Deep Learning and Attention Encoder Decoder Recurrent Neural Networks (RNN). NLP is used for parsing, reduction of words and to generate text summary in abstractive summarization. However at present NLP is a low cost technique and lacks precision. We propose a new dataset consisting of a select bunch of research papers to ensure the correctness of the system.**

Keywords— Neural Networks, Text, Summarization, NLP, Deep Learning, Encoder Decoder

I. INTRODUCTIN

Information and digital data in particular has expanded exponentially in the last few years. The world saw a dotcom boom at the turn of this century and it is estimated that 90% of the data has been created in the last 2 years alone. This increase in information has also led to a need to obtain a summarized version of data that can convey the basic crux of a situation in fewer words. With time becoming a rare commodity, individuals are constantly looking to acquire more in less time. This ideology formed the basis of our topic selection – to convey more by using less. Data Summarization as a concept has been cherished right from the inception of data not just digitally but also in ancient times for written content. Summaries were often substituted for the original document whenever time was a restricting parameter. However, the amount of information produced digitally has eclipsed all the other forms of information combined. This has led to summarization becoming more essential than ever before.

II. HISTORY

A. Automatic Summarization

Automatic summarization is the process of shortening a text document with software, in order to create a summary with the major points of the original document. Technologies that can make a coherent summary take into account variables such as length, writing style and syntax.

Automatic data summarization is part of machine Learning and data mining. The main idea of summarization is to find a subset of data which contains the "information" of the entire set. Such techniques are widely used in industry today. Search engines are an example; others include summarization of documents, image collections and videos. Document summarization tries to create a representative summary or abstract of the entire document, by finding the most informative sentences, while in image summarization the system finds the most representative and important (i.e. salient) images.[Citation needed] For surveillance videos, one might want to extract the important events from the uneventful context.

Most early work on single-document summarization focused on technical documents. Perhaps the most cited paper on summarization is that of (Luhn, 1958), that describes
Research done at IBM in the 1950s. In his work, Luhn proposed that the frequency of a particular word in an article provides an useful measure of its significance. There are several key ideas put forward in this paper that have assumed importance in later work on summarization. As a first step, words were stemmed to their root forms, and stop words
Were deleted. Luhn then compiled a list of content words sorted by decreasing frequency, the index providing a significance measure of the word. On a sentence level, a significance factor was derived that reflects the number of occurrences of significant words within a sentence, and the linear distance between them due to the intervention of non- significant words. All sentences are ranked in order of their significance factor, and the top ranking sentences are finally selected to form the auto-abstract.

Edmondson (1969) describes a system that produces document extracts. His primary contribution was the development of a typical structure for an extractive summarization experiment. At first, the author developed a protocol for creating manual extracts that was applied in a set of 400 technical documents. The two features of word frequency and positional importance were incorporated from the previous two works. Two other features were used: the presence of cue words (presence of words like significant, or hardly), and the skeleton of the document (whether the sentence is a title or heading). Weights were attached to each of these features manually to score each sentence. During evaluation, it was found that about 44% of the auto-extracts matched the manual extracts

B. Machine Learning Models

In the 1990s, with the advent of machine learning techniques in NLP, a series of seminal publications appeared that employed statistical techniques to produce document extracts. While initially most systems assumed feature independence and relied on naive-Bayes methods, others have focused on the choice of appropriate features and 3 on learning algorithms that make no independence assumptions. Other significant approaches involved hidden Markov models and log-linear models to improve extractive summarization. A very recent paper, in Contrast, used neural networks and third party features (like common words in search engine queries) to improve purely extractive single document summarization.

The methods and models used were –

1) *Naive-Bayes Methods*
2) *Rich Features and Decision Tree*
3) *Hidden Markov Models*
4) *Neural Networks*

C. Mult- Document Summarization

Extraction of a single summary from multiple documents has gained interest since mid-1990s, most applications being in the domain of news articles. Several Web-based news clustering systems were inspired by research on multi-document summarization, for example Google News, Columbia News Blaster, News In Essence. This departs from single-document summarization since the problem involves multiple sources of information that overlap and supplement each other, being contradictory at occasions. So the key tasks are not only identifying and coping with redundancy across documents, but also recognizing novelty and ensuring that the final summary is both coherent and complete.

D. Problem Definition

The primary problem definition in our project is about extracting key information from research papers by using a text summarization tool. It involves the objective of reading the input data in the form of articles or documents which are analyzed and preprocessed in order to split it into several segments that are each attributed values. These are then converted into concrete semantic statements using several algorithms which will be mentioned. Our problem definition covers the concept of converting convoluted, mammoth data into a useful segment which is concise, yet does not lose the essence of the original document.

The project was conceptualized keeping in mind the fundamental goal of creating a tool which is able to convert towering amounts of data into concise summaries which have the potential to not just save time but also help reduce the amount of space required to save the data. File sizes can decrease by a huge percentage if the essence of the article is maintained and the paper has a lossless summary derived. While there are several milestones that we have determined, some of them are – obtaining keywords from articles and papers, creating a summary without diving into specifics, creating a summary that contains the specifics as well, obtaining semantically correct outputs which are human readable in nature.

The scope of this project is obtaining a tool which can efficiently translate data into a summarized form suitable for human readers. It involves detecting and extracting key words which can be used in generating the summary. Word groups based on glove values using RNN, LSTM and training our data set forms the basis of our implementation. These are useful to generate both extractive summaries and abstraction based summaries which will help to simplify major documents; in our case, that refers to the research papers and articles that we will use as a data set. Our end product deliverable will be a summary that will be lossless in terms of quality of information and be efficient enough to contain all the specifics of the original document. With no hardware requirements, the primary tool will be created and trained using existing libraries and frameworks by working on available workspaces like Google Colab

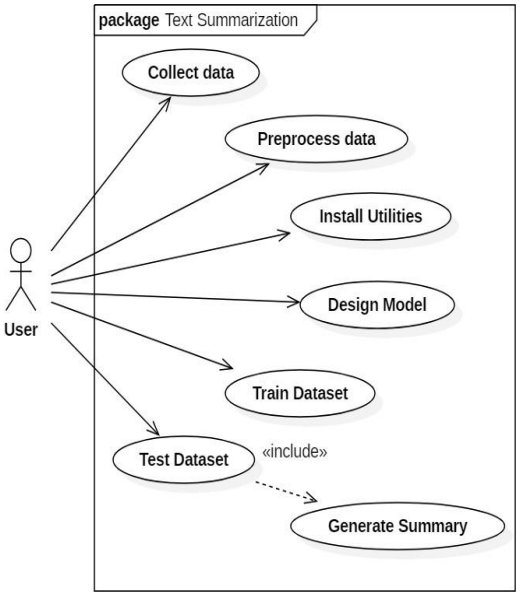
The project has a lot of scope of becoming something revolutionary, purely because of the sheer application that it incorporates due to the constantly booming data sources. While we intend to create a tool that will primarily focus on research papers and articles within a particular domain or realm, it is highly possible to expand the scope of the project to a wide array of applications. News articles have been converted into smaller bytes by In Shorts which shows the significance of summarization. Market trends, sentiment analysis, political inclinations and personal data points on individuals can all be summarized to provide a detailed yet concise output which can be stored in a much more controlled manner, using much lesser space. As someone looking to venture in the realm of Data Science, this concept holds widespread applications because it can assist in improving the process of preprocessing. Relevant keywords and data summaries can help eliminate the garbage values which will in turn speed up the process of extracting useful insights from available information..

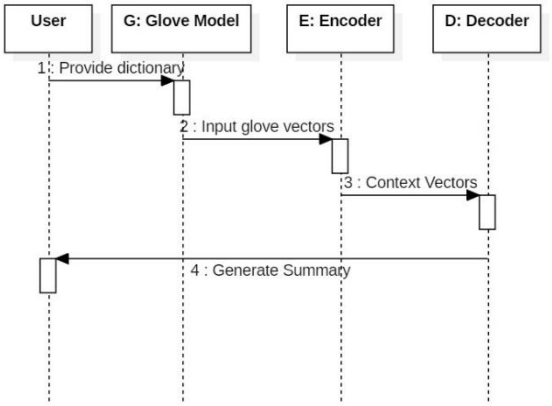
E. Proposed Aspects of the Project

This report is built to generate implementation approaches to address text summarization. It is built to simply run on Google Colab, in one notebook so one would only need an internet connection to run these examples without the need to have a powerful machine. The model generated is being trained using multiple data sets which have the potential to create a refined tool that can implicitly generate lossless summaries while keeping the crux and specifics of the information intact.

A seq2seq network with various rich features is used to begin
with the implementation. It's a machine translation model that was introduced by Google. Using deep learning, the model not just takes the current word into account but also its neighborhood. It takes as input a sequence of words and generates an output sequence of words. It does so by use of the recurrent neural network (RNN). Although the vanilla version of RNN is rarely used, its more advanced version i.e. LSTM or GRU are used. The concept of a pointer generator

is introduced to solve some flaws prevailing within the seq2seq network

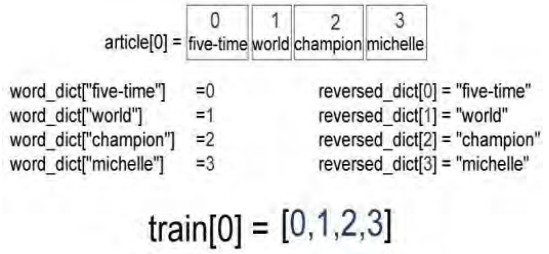




III. .SYSTEM ANALYSIS

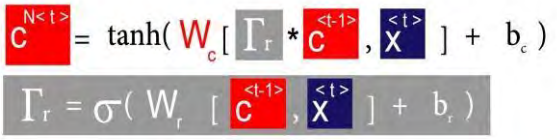
A. Techniques Used

*1) **Word Embedding** – For the text summarization to work, you must represent your words in a dictionary format. Each word in the corpus should have a representation in the dictionary. We would need the reverse operation also. After building the dictionary for our data, we would begin to build the actual dataset that would be used in our algorithm. The dataset would be represented by simply getting the collection of word dictionary for the words in the given sentence.*



*2) **Seq2seq** - RNN has a many-to-many architecture with same lengths for both input and output. However for text summarization, we need to have the ability to have different lengths for input and output. We need a special network that takes input of length (Tx), and generates another output of another different length (Ty), this architecture is called Encoder Decoder. Both the Encoder- Decoder here are RNN network, but encoder uses input, and generates an output state that is then used as input to decoder stage. However, there are 2 main problems with the RNN unit namely exploding gradients and vanishing gradients. Both GRU & LSTM solves the problem of vanishing gradients that normal RNN unit suffers from, they do it by implementing a memory cell within their network , this enables them to store data from early within the sequence to be used later within the sequence.*

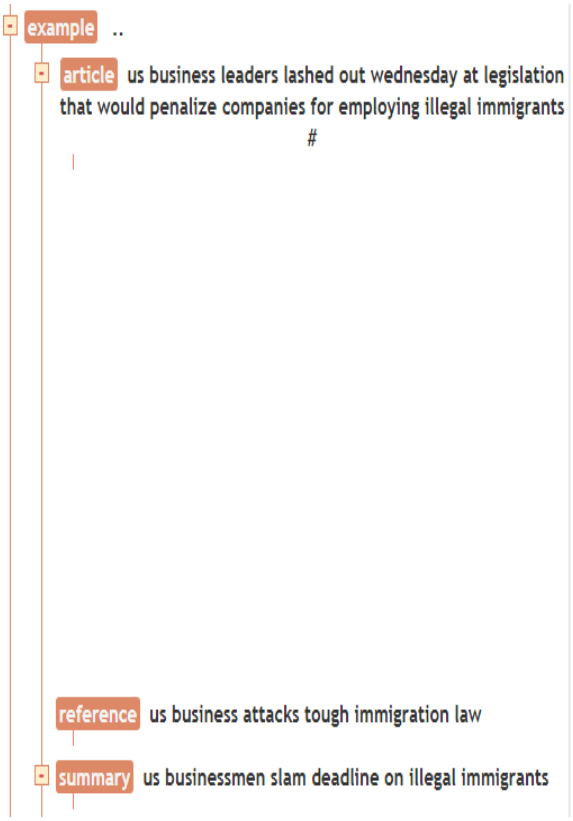
*3) **Gated Recurrent Unit (GRU)** - We have 4 main equations that govern the GRU:*



C here denotes for the memory cell , here it would be the output of the GRU cell. The N sub letter denotes that it is the newly proposed C value (we would use it latter to generate the real c output of the GRU. So here the new proposed output C (candidate), would depend on the old output C (old candidate) , and the current input at that time. To remember the value of C (candidate), we use another parameter called F (gate update) , this would control whether or not we would update the value of C via a sigmoid function. We also have a learnable (Fr) parameter to learn the relevance between C new and C old.

B. Algorithm

C. User Interface Design(Snapshot of Result)



IV. CONCLUSION

We have successfully been able to work on several aspects of creating a tool for text summarization. The state of the art has been found to be efficient, but this is the case of abstractive summarization methods, with the same data that we utilized. Of the literature reviewed, we found that the
best-performing methods extractive methods close to 11%.

Understandably, it is difficult to reach high values with extractive methods. We created and used algorithms to find and break the sentences to words that would generate the best summaries. Glove values have been used above while we utilized the concept of encoder-decoder to obtain words that can be related. We believe that such a metric should be "mandatory" on any paper proposing an extractive summarization method, as it sets a good reference - the direct opposite of the random baseline. The seq2seq method is implemented in its entirety but is improved on by using modern methods to increase efficiency.

The interesting thing is the variation, showing that summaries vary in their use of original words. By changing the methods by which we are generating summaries and training the models, we will be generating varied summaries for the same data set.

V. ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R.B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

[1] H. P. Luhn. The automatic creation of literature abstracts. IBM Journal of Research and Development, 1958.
[2] H. P. Edmundson. New Methods in Automatic Extracting. Journal of the Association for Computing Machinery, 1969.
[3] David M. Blei. Probabilistic Topic Models. Communications of the ACM, 2012.
[4] Alexander M. Rush, Sumit Chopra, Jason Weston. A Neural Attention Model for Abstractive Sentence Summarization. EMNLP, 2015.
[5] Siddhartha Banerjee, Prasenjit Mitra, Kazunari Sugiyama. Generating Abstractive Summaries from Meeting Transcripts. arXiv:1609.07033, Proceedings of the 2015 ACM Symposium
[6] Federico Barrios, Federico López, Luis Argerich and Rosa Wachenchauzer. Variations of the Similarity Function of TextRank for Automated Summarization. arXiv:1602.03606, 2016.
[7] Suzuki, Jun, and Masaaki Nagata. Cutting-off Redundant Repeating Generations for Neural Abstractive Summarization. EACL 2017 (2017): 291.
[8] Shibhansh Dohare, Harish Karnick. Text Summarization using Abstract Meaning Representation. arXiv:1706.01678, 2017.
[9] André Cibils, Claudiu Musat, Andreea Hossman, Michael Baeriswyl. Diverse Beam Search for Increased Novelty in Abstractive Summarization. arXiv:1802.01457, 2018.
[10] https://www.researchgate.net/publication/325836782_Query-oriented_text_summarization_using_sentence_extraction_technique(02/10/2019).
[11] https://www.coursera.org/lecture/nlp-sequence-models/gated-recurrent-unit-gru-agZiL(23/09/2019).
[12] https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network(20/09/2019).
[13] https://colah.github.io/posts/2015-08-Understanding-LSTMs/(15/10/2019).

SYSTEM TO PREDICT STOCK PRICES USING ML AND NLP

Ashish Kothari , Sumit Joshi , Ved Khetle , Mrs.ShiwaniGupta(Deputy HOD)
Dept. ofComputer Engineering
ThakurCollege ofEngineeringand Technology
ashishkothari056@gmail.com , sumit.jsvp@gmail.com , shiwani.gupta@thakureducation.org

Abstract—**Stock priceforecastingisapopularandimportant topicinfinancialandacademicstudies. Itisanunreliable placefor prediction,sincethereare nosignificantrulesto estimateorpredictthe priceofshareinthesharemarket.Also theopinion of peopleworldwideaboutaparticularcompany haseffectsonitsstockprice.Negativeopinionsontheweb maycontributetoa downward trend in thestock price&positive feedbackincreases it. Many methods like technical analysis,fundamental analysis,timeseries analysisand statistical analysis,etc. areusedtoattempttopredictpricein thesharemarketbutnone of thesemethods are provenas consistentlyacceptable predictiontoolsincethey arehighly unreliable.The method used inthispaperempowers an inexperiencedindividualtoenter the stockmarketwithaboost inconfidenceandmaketradesonthestockmarket.It provides reliable analysis basedonmultiplemathematicalcomputational algorithmsto an individual tofallbackuponif andwhen needbe.Inthis paper we attempttobuild a Machine learning modelto predictthe stockmarketprices. Auto ARIMA model has beeneffectivelyimplementedforforecastingstock prices,returns, andstockmodeling. Thisworkis beneficialfor anytrader as predictionisdone onthe listedcompaniesStock Exchange.Weselectacertaingroup of parameterswith relativelysignificantimpacton theshare priceofacompany. Therelationbetweentheselectedfactors andsharepriceis formulatedwhichcan helpinforecastingaccurate result.But eventhenthe predictionmay have some inaccuraciesandcan be improvedin accuracy.We aim toincrease the accuracy of ourpredictionbyalsoanalyzingtheopinionsandsentiment s of other people onthe internet.**

Keywords—AutoARIMA,time series forecasting,opinion mining, Stockmarketintelligence.

I. INTRODUCTION

Thefieldoffinancial modelling ismorematurethan one might perceive. Moreso,whenthefactthatallthe majormodellingandforecasting algorithms wereinitially employedinthisdomain. Thepursuitofhaving successfullymodelledthepathofatickerwithperceptible accuracygaverisetoageneration ofresearchers and developers, whoassociatethemselveswiththeterm Quant'.

WhataQuantdoesisitgeneratesfinancialmodels [3] andteststhemagainstthemarkets fortheiraccuracyin facultiesliketrajectoryprediction, priceprediction, risk assessment,varianceandconsumptiontonameafew.This

Pursuitlettoatremendous increaseinresearchinfinancial domainwiththesolepurposeofharnessing thechaos;for monetarybenefitsandothers.

Tosupportthisendeavormultiplealgorithmshavebeen developed inthespanofafewdecades, andtheemergence ofeasy toimplementlibrariesinlanguageslike Pythonhave madeitwidespread intruesense.Thefirstofthealgorithms tobe usedforforecasting thefuturetrends ofstockprices wasthecrudelinear regression. Linearregressionisthe analysisoftwoseparatevariables, usedtodefineasingle relationshipandisausefulmeasure fortechnical and quantitative analysisinfinancial markets.Plottingstock pricesalonganormal distributioncanallowtraderstosee whenastockisoverbought oroversold.Thisalgorithm, whenusedeffectively,allowedtraderstostrategically place keypricingpointswhichisentry,exitandstop-lossprices.

Butastheevolutionwasspearheaded byconstantly evolvingalgorithms,therewasaproliferation of sophisticated, multifactoralgorithms,likeQ-learning, Boost,Last Value,etc., that providedabetterandholistic view when the nature of the market isconcerned.This pseudoevolution furtherledtoconvergenceofanother parallel domainofNLP.Withtheriseinopinionavailable overtheinternet,acrudebiasproves beneficial in determiningthegeneralsentiment ofmassesregardingany particularcompanyalsoNLPhelpsintrendanalysis.

Inrecenttimes,multiplefinancialorganizations and hedge fundshaveimplemented intelligent systems toperformHVTand forecastingbeforemaking a position. Theyalsoemploytheseprowessto provide serviceslikerisk analysisandassetmanagement.

Theadventofcomputational powerandthe commonplacenatureoftechnicalskillsledtotheriseinthe fieldoftechnological applicationinfinances.Wehere compareandtrytocombinethesealgorithmsanddisciplines togetcomparativelybetterresultsinstockpriceforecasting domain.

II. LITERATUREREVIEW

Ye	Author	Affiliation	MainObjective	GapIdentified
20 17	RafekTicki n	Kingston,ON,CA.	Forecastingdiffusionofnew technologies.	consistent data adversely affects the output.
20 14	Debadrita Banerjee	ers College KOL, IN.	Develop appropriate model to forecast future unobservedvaluesofIndianstockmarke	This modelmay notbeabletocapturethee5ffect ofeconomic variables.
20 17	Lakshmi Yermaletal	Amrita Coimbatore,IN	ofAutoArimausingEviews9.5for forecastingstock data	Negligence towardsimpactof corporate
20 18	ShriBharath etal	B.SAbdurRahman Unversity, TN,IN	mpact of useropinionon Sensexusing lexiconsentimentclassification.	Only uses opinions fromRssfeed andnotFacebookandTwitter.
20 15	Aditya Bhardwajet al	NITTTR CH,IN	Extractthe subjective contentby analyzing user's opinion	Noclear distinction amongst varioustextsentiment
20 18	Dev Shah etal	Kingston,ON,CA.	Developmentofsentimentanalysis Dictionaryforfinancial sector	Discard volatilityofstockwith minimalthreshold of2% Price change.

A.RegressionBasedForecasting.

Regressionmodelisapredictiveequationwhichrelates a responsevariableordependentvariable(Y_t),say,sales ofamerchandise,toasetofindependent variables(X_t, Z_t,etc.),say,priceofthemerchandise andGDP.The subscript,t,denotesatimesequence; Y_T, $X_t$and$Z_t$are time seriesdata. If Ytis driven by Xtand Tzu, aregression modelcanbebuilttopredicttheresponse variable(Y_t),giventhevaluesofindependent variables ($X_t$and Z_t).This type of modelingwillrequireahistoric database fortheresponsevariable(Y_t)andindependent variables(X_t andZ_t).Theregressionmethodcanbe usedtomodelanysituation aslongasthereexistsa strongcausalrelationship betweenthedependent and independent variables.Forbusinessforecasting, regression modelisthefirsttobetried,andinmany casesregressionmodelsaresufficient.Theoretically, a goodregression modelrequires notonlythatitpasses variousstatistical testsbutalsoitupholdsallthe underlyingassumptions.Inrealworldsituations,one canseldomdevelopaforecastingmodelthat satisfiesall theconditions. Amodelwhichdoesnotsatisfyallthe conditions maybestillusefultoproduceaforecastthat meetstheuser'srequirements.

B.Q-LearningbasedForecasting.

Inthelastfew decadesforecastingofstock returnshas becomeanimportant fieldofresearch.Inmostofthe cases theresearchers had attemptedtoestablisha linearrelationship betweentheinputmacroeconomic variables andthestockreturns.Afterthediscovery of nonlinearity inthestockmarketindexreturns,many literatureshavecomeupinnonlinearstatistical Modelingofthestockreturns,mostofthemrequired That the nonlinear—modelbe specifiedbeforethe estimationisdone.Butsincestockmarket returnis noisy,uncertain,chaoticandnonlinearinnature,ANN hasevolved outtobebettertechniqueincapturing the structural relationship betweenastock'sperformance anditsdeterminantfactorsmoreaccuratelythanmany otherstatisticaltechniques. Inliterature, differentsets ofinputvariables areusedtopredictstockreturns.In fact,differentinputvariables areusedtopredictthe samesetofstockreturndata.Someresearchers used input datafromasingletimeseries where others consideredtheinclusionofheterogeneous market informationandmacro-economic variables.Some researchers evenpre-processed theseinputdatasets beforefeedingittotheANNforforecasting.

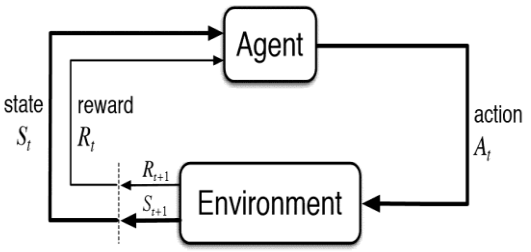


Fig 1: Q-LearningBased Forecasting

C.NaturalLanguageProcessing

Sentiments of themassesthatmakeupthemarketplayjustasimportan t aroleindefining the priceofanindexasacompany'sfundamentals do. Thus, appropriately analyzing theirsentimentsvia NLP[9]playa crucialroleindecidingthe successrate oftrendprediction.

NLPviaANN:AnartificialneuralnetworK(ANN)[4] is acomputational nonlinear modelbasedontheneuralstructureofthebrainthatisabl etolearnto Performtaskslikeclassification, prediction,decision- making,visualization,

andothersjustbyconsidering examples.Anartificial neuralnetwork consists of artificial neuronsorprocessingelements andis organizedinthreeinterconnectedlayers: input, hidden thatmayincludemorethanonelayer,andoutput.

III. PROPOSEDMETHODOLOGY

A. MethodforPrediction

ARIMAforecasting

UsingARIMA model,youcanforecast atimeseries [2]usingtheseries past values. Inthispost,webuild anoptimalARIMA modelfromscratchandextendit toSeasonalARIMA(SARIMA) andSARIMAX models. YouwillalsoseehowtobuildautoARIMA models inpython[3].Atimeseriesisasequence where a metric is recorded over regular time intervals.Depending onthefrequency, atimeseries canbeofyearly(ex:annualbudget), quarterly (ex: expenses), monthly (ex:airtraffic),weekly (ex:sales qty),daily(ex:weather), hourly(ex:stock'sprice), minutes (ex:inbound callsinacallcanter)andeven secondswise (ex: web traffic). ARIMA,short for

'AutoRegressive IntegratedMovingAverage'is actually aclass ofmodelsthat 'explains'agiventime seriesbasedonitsownpastvalues,thatis,itsown lagsandthelaggedforecast errors,sothatequation canbeusedtoforecastfuturevalues.

Any'non-seasonal' timeseriesthatexhibitspatterns andisnotarandom white noisecanbemodeled with ARIMA models.AnARIMAmodelischaracterized by3 terms:p, d,q where,

p istheorderoftheARterm q istheorderoftheMAterm

disthenumberofdifferencingrequired tomakethe timeseriesstationary

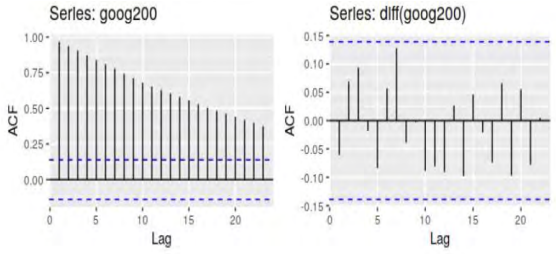


Fig.1: Differencing onTimeSeries

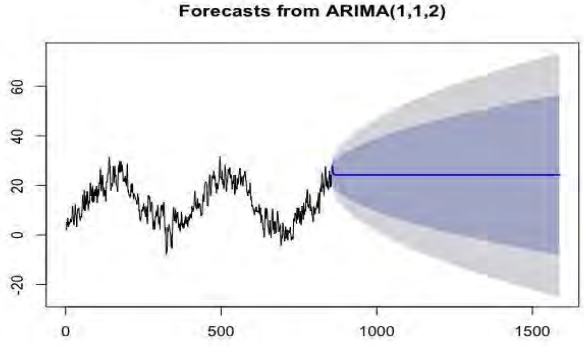


Fig3: ARIMAPlot

Ifatimeserieshasseasonal patterns, thenyouneedto addseasonaltermsanditbecomesSARIMA,shortfor 'SeasonalARIMA'.

B. NLP

Sentiment analysisiscarriedoutontwitterplatformin relationtothecompany whichispubliclytraded.The PythonlibraryTweedy isutilizedforperforming sentiment analysis,whereas thetweetsareobtainedfrom TwitterAPI.Tweedyworksinthefollowingway:

1. TweetScraping
2. Tokenization
3. PoSTagging
4. SemanticAnalysis
5. CorporaComparison
6. Results

Tweedysplitsthe scrapedtweetsin individualwords calledtokens,thenthesetokensaretaggedaccording to partof speechandanalyzedsemantically. Later,thesepre-processed tokensarecomparedtoword corpora[1]andthe sentimentsare determined.

Sentiment analysis isbasicallytheprocessofdeterminingtheattitudeorthe emotionofthewriter,i.e., whetheritispositiveornegativeorneutral. The*sentiment*function of textblobreturns two properties,polarity, andsubjectivity.Polarityis float whichliesintherangeof [-1,1]where1means positive statement and-1means anegativestatement. Subjective sentences generallyrefertopersonalopinion,emotion or judgmentwhereasobjective referstofactualinformation. Subjectivity isalsoafloatwhichliesintherangeof [0,1].The predicted pricesareassociatedwiththe confidencelevels, and the range fluctuationfrom the actualpricefollowsaninversely proportional relationship.Anotheraspecttoagoodpredictionis checkingif thesentimentsof the massesindicate thesame.Here,sentiment analysis[4][5]comesinto picture.Currently,thisfeatureworksonTwitter,using the officialTwitterAPI.Thescrapingstartsforanytweet thatcontainsthekeywordandtheretrievedtweetsarefed tothePythonlibraryTweepy, wherefurtherprocessing happens.

```
Enter the tag to analyse : Tesla
Positive tweets percentage: 31.11111111111111 %
Negative tweets percentage: 6.666666666666667 %
Neutral tweets percentage: 62.22222222222222 % \
```

Fig 4: SentimentMetric

Here as showninfig4above,thetaggedword is 'Tesla',and theAPIscrapesalltherecent tweetsandperformssentiment

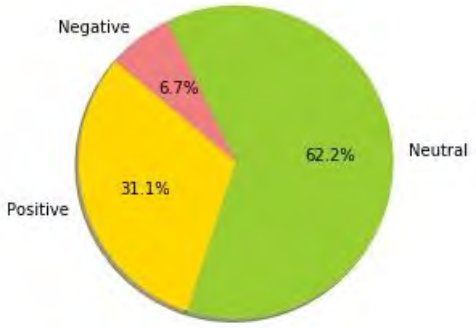


Fig 5: SentimentBias

Thisasshowninfig5abovegivesabetter view regarding outlook foranycompanyandservesaseitheranaffirmation orprobableadmonishment forthepredict ion.An amalgamationof all these, incorporatedwith fuzzy rules, wouldgiveasuitableoutlookandtrajectorypredictionfor anystock.Thefuzzyrulesaretobeimbibedonthemetric of sentiment. The majorsentiment is to be assigned higher value,andistobetheaffecting factoronstockprice accordingly.

IV. RESULTS ANDDISCUSSION

Thealgorithmindividuallyperformsat 98%accuracy consistently, butonlyonshorttimespans.Whenalonger durationisconcerned,thefactors,mostlyunknown, increase exponentially withincreasingtimeframe.Thus,adynamic, iterativeforecastthattakesintoconsideration allthecurrent happeningsalongwiththemasssentimentisnecessary.

Dynamic, continuousforecasting and relentless monitoringofcurrent happenings ismustforavoidingany potentiallosses.

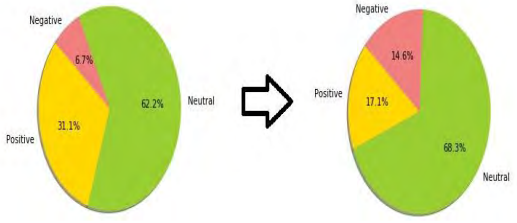


Fig 6: Changein resultswithMultipleCorpora

Thebelowfig7showshowutilizingmultiplecorporasfor analysingsentimentsleadstoa better,moreaccurateresult. ThedashboardGUIacceptstheticker,i.e.thelistingname ofthe stock,andprovidesseasonalaswellasnon-seasonal pricepredictionsasfollows.

analysisonthem.Thetoptweetsforanysentimentcanalso beretrievedforfurtheranalysisandreview.

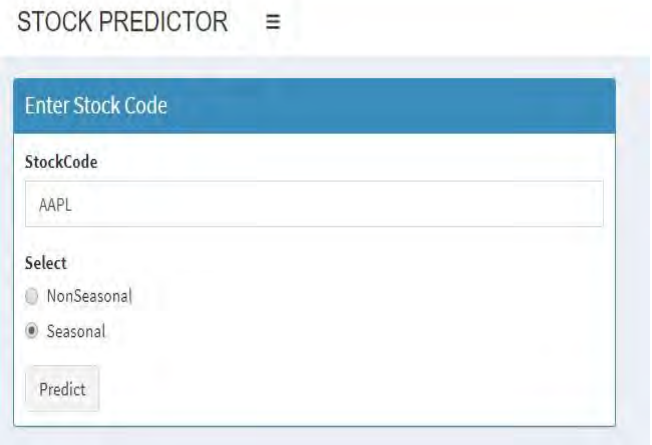


Fig 7: Dashboard

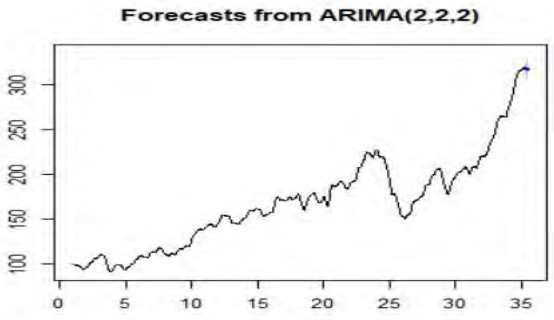


Fig 8: SeasonalAuto ARIMA Forecast

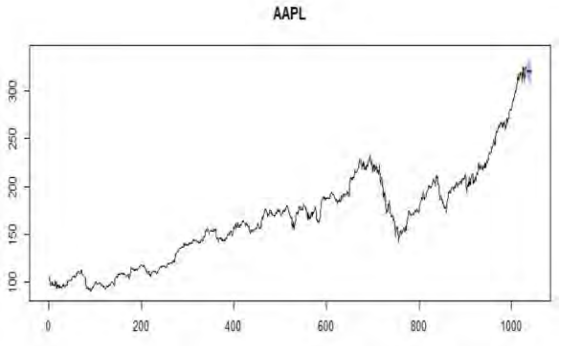


Fig 9: Nonseasonal Auto ARIMAForecast

Thefigures8 and9 aboveshowsthegraphofstock pricesofApple(AAPL)listedonNYSE.Figure8 showsthe predictionwiththeseasonalcomponentconsidered.Figu re9 showstheprediciton withotconsideringthe seasonal component.

Auto Arima - Non Seasonal

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
640.81	629.59	652.03	623.65	657.97
640.81	624.94	656.68	616.54	665.08
640.81	621.37	660.25	611.08	670.54
640.81	618.36	663.26	606.48	675.14
640.81	615.71	665.91	602.43	679.19
640.81	613.32	668.30	598.77	682.85
640.81	611.12	670.50	595.40	686.22
640.81	609.07	672.55	592.26	689.36
640.81	607.14	674.48	589.32	692.30
640.81	605.32	676.30	586.53	695.09

Fig 10: Non- SeasonalPriceprediction

V. CONCLUSION

Thispaperaimstopredict different marketinvestment strategies available fortheinvestors.Byprovidingcrisp decision basedonpredicted futurestock value.The AutoArimamodelalongwith thesentimentanalysis will maketheinvesting processclosetoperfection andflexible. Weanalyzedtherequirements andobjectivesofourproject. We createdbasic block diagram representing the functionality ofourprojectalso,weanalyzedsoftwaretobe usedforefficientexecution. Wecameacrossvarious algorithmsand datasetswhich will be beneficialfor our project.Wearecurrently workingonidentifying more algorithms suitableforourproject.Wearealsoworkingon dividingourprojectintomodulesandworkingcollectivel y oneachmodule.Thesemoduleswillbethebasicbuilding Blocksofourprojectwhichwillbecodedandintegrated togetherlater.

VI. FUTURESCOPE

Withtimeandincreasingsophisticationofalgorithms,th e functionalityofassessingthefundamentalsofthe organizationasa wholebasedontheQoQresultscanalsobe added,whichcurrentlyreliesonlyonhumanexpertise. Additionally,HVT(HighVolumeTrading)canalsobe incorporatedwiththecomputationtimesreachingreal-time speeds.

References

[1] RafekTickin, “Analysis of multiple Corpora and their effect on sentimentanalysis accuracy”,Worldconferenceontech,132-648, ABDH,2017

[2] DebadritaBanerjee, “Forecastingof IndianStock(MarketusingTime-seriesARIMAModel”,ICBIM,KOL,June 2014

[3] LakshmiYermalet al,“Application ofAutoARIMAModelfor ForecastingReturnsonMinuteWiseAmalgamated DatainNSE”,IEEEInternational Conference on ComputationalIntelligence and ComputingResearch, Nov2017

[4] ShriBharathietal,“SentimentAnalysisforEffectiveStockMarket Predion”,IEEE,IN,June 2017

[5] AdityaBhardwajetal,“SentimentAnalysisforIndianStockMarket PredictionUsingSensexandNifty”,4th InternationalConference on Eco-friendlyComputing and Communication Systems, CH,Dec 2015

[6] DevShahetal,“PredictingtheEffectsofNewsSentimentsonthe Stock Market”,IEEE,CA,Nov2018

[7]https://www.machinelearningplus.com/time-series/arima-model-time- series-forecasting-python/(23/10/19)

[8] https://medium.com/@datamonsters/artificial-neural-networks-for- natural-language-processing-part-1-64ca9ebfa3b2 (10/10/19)

[9] https://courses.analyticsvidhya.com/courses/natural-language-processing-nlp,(20/7/19)

[10] https://www.machinelearningplus.com/time-series/arima-model-time- series-forecasting-python/, (30/8/19)

Design and development of descriptive answer assessment using NLP

Karan Agarwal , Mukul Garud , Sandesh Gade
Computer Engineering, TCET
University of Mumbai
Karan.agarwal1799@gmail.com , Mukulgarud9@gmail.com , Sandeshgade21@gmail.com

Abstract **The following section provides an overview of the derived web application for the descriptive online examination system. The current way of checking subjective paper is adverse. Evaluating the Subjective Answers is a critical task to perform. When human being evaluates anything, the quality of evaluation may vary along with the emotions of Person. In Machine Learning, all result is only based on the input data provided by the user.To begin with the purpose of the system is to conduct online descriptive exam in an efficient manner in terms of cost, time and safety, the system also tend to provide the area of improvements required by both teacher and student. The system would use the Natural Language Processing (NLP) for verification of answers and would use web development technology for the design of the web application. We are trying to extract meaning from the answer so that online assessment can be more accurate.**

Keywords: Tokenization, lemmatization,Stop words, bags-of-words, part of speech(POS), WordNet.

I. INTRODUCTION

Online Descriptive answer Assessment, an online tool that tackles your test requirements with friendly interfaces. It is mainly focused on managing university/school type exams containing descriptive questions or a mix of descriptive and objective type questions.

The current way of checking subjective paper is adverse. Evaluating the Subjective Answers is a critical task to perform. When human being evaluates anything, the quality of evaluation may vary along with the emotions of Person. In Machine Learning, all result is only based on the input data provided by the user.

The following section provides an overview of the derived web-application for the subject online examination system. To begin with the purpose of the document is presented and is intended to the audience. Subsequently, the scope of the project specified by the document is given particular focus on examination system and its relevant benefits. To conclude, a document overview is provided to felicitate better examination process**.**

.

II. LITERATURE SURVEY: -

Thomas N. T. Ashwini Kumar and Kamal Bijlani implemented e-learning system like Learning Management System (LMS) to assess candidate's descriptive answers and provide immediate feedback to teachers.[5] This is implemented by comparing candidate's answers with teacher's ideal set of answers using latent semantic analysis. Then evaluating order of previous and upcoming words in the answer using positional indexing based on the keyword list added by the teacher. Then final score is generated.

C rater [4] is a technique used for the content scoring. This content scoring is based on model building which makes various answer model for candidate's short answer. ETS (Education Testing Service) is used for examining small answer of candidate having near about 100 words. It uses systematic method & rubrics element which specify accurate & significant terms which should be exist in candidate's answer. The problem of dissimilar answer contents of candidate is resolved using c-rater. C rates includes the steps like Model building, C Rater automatically processes, maintaining algorithm gold map, Apply candidate answer**.**

III. TECHNOLOGY USED

NLP: - Natural Language processing is a wide domain covering concepts of Computer Science, Artificial Intelligence and Machine Learning. It is used to analyze text or how humans speak. One of the applications of NLP is Semantic Analysis(Understanding the meaning of text).

Django: - Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

React JS: -Reactis a JavaScript library for building user interfaces. React can be used as a base in the development of single-page or mobile applications, as it is optimal for fetching rapidly changing data that needs to be recorded. However, fetching data is only the beginning of what happens on a web page, which is why complex React applications usually require the use of additional libraries for state management, routing, and interaction

with an API: Redux' React Router and axiosare examples of such libraries.

IV. IMPLEMENTATION DETAIL

Several methods have been suggested for descriptive answer assessment. The approaches are mainly based on text mining technique which involves keyword matching, sequence matching and quantitative analysis and semantic analysis.

- The system will use herically organized oracle system for assessment.
- System will use NLP to generate meaning for assessment.
- The system will be useful for subjective questions.
- The accuracy is high due to meaning extraction.

There are many ways text analytics can be implemented depending on the business needs, data types, and data sources. All share four key steps.

Step 1: Data Acquisition:

Text analytics begins with collecting the text to be analyzed -- defining, selecting, acquiring, and storing raw data. This data can include text documents, web pages (blogs, news, etc.), and online reviews, among other sources. Data sources can be internal or external to an organization.

Step 2: Data Preparation:

i. Once data is acquired, the enterprise must prepare it for analysis. The data must be in the proper form to work with machine learning models that will be used for data analysis. There are four stages in data preparation:

ii. Text cleansing removes any unnecessary or unwanted information, such as ads from web pages. Text data is restructured to ensure data can be read the same way across the system and to improve data integrity (also known as "text normalization").

iii. Tokenization breaks up a sequence of strings into pieces (such as words, keywords, phrases, symbols, and other elements) called tokens. Semantically meaningful pieces (such as words) will be used for analysis.

iv. Part-of-speech tagging (also referred as "PoS") assigns a grammatical category to the identified tokens. Familiar grammatical categories include noun, verb, adjective, and adverb.

v. Parsing creates syntactic structures from the text based on the tokens and PoS models. Parsing algorithms consider the text's grammar for syntactic structuring. Sentences with the same meaning but different grammatical structures will result in different syntactic structures.

Step 3: Data Analysis:

Data analysis is the process of analyzing the prepared text data. Machine learning models can be used to analyze huge volumes of data, and the outcome is typically produced as an API in JSON format or in a CSV/Excel file. There are many ways data can be analyzed, two popular approaches are text extraction and text tagging.

Text Extraction is the process of identifying structured information from unstructured text. Text Tagging is the process of assigning tags to text data based on its content and relevance.
Two common models for text tagging are "bag of words" and "Word2vec."

The bag-of-words method is the easiest method to understand, but it's outdated and has been deprecated. This method simply counts the number of words within the text content regardless of location and context. The disadvantage of this technique is that it does not offer a way to understand context from words content with a higher word count is given a higher (and, falsely, more relevant) score.

Word2Vec has become the preferred method of text tagging. Text collected for Word2Vec is turned into a vector, which provides relevant information about words (including synonyms). For example, the terms "man" and "boy" can be closely related. Word2Vec also understands that the words "humor" and "humour" should be treated the same way. Word2Vec produces a mesh of related words. The closer the words are to each other in the neural network, the stronger their relationship to each other. This neural net allows algorithms to better understand the context of words, so data scientists can generate better analysis of content relevancy.

We are checking the similarity between two answers using some NLP methods such as synsets(), wup_similarity(),etc. Synsets are organized in a Hypernym tree. This tree can be used for reasoning about the similarity between the Synsets it contains. The
closer the two Synsets are in the tree, the more similar they are. The wup_similarity() method is short for Wu-Palmer Similarity, which is a scoring method based on how similar the word senses are and where the Synsets occur relative to each other in the hypernym tree.

Step 4: Data Visualization:

Visualization is the process of transforming

$$\text{Wu} - \text{Palmer} = 2 * \frac{\text{depth}\,(\text{lcs}(s1, s2))}{(\text{depth}\,(s1) + \text{depth}\,(s2))}$$

analysis into actionable insights, representing the data in graphs, tables, and other easy-to-understand representations. Organizations can use a wide variety of commercial and open source visualization tools.

The Role of Natural Language Processing:

NLP is a component of text analytics. Most advanced text analytics platforms and products use NLP algorithms for linguistic (language-driven) analysis that helps machines read text. NLP analyzes words for relevancy, including related words that should be considered equivalent, even if they are expressed differently (e.g., humor vs. humour). It's the workhorse behind steps 2 and 3 described above.

One popular application of NLP is identifying relevant, quality content for search engines. For example, Google uses NLP in several ways, the most prominent of which is in search engine organization and categorization. Long ago, a webmaster could achieve a higher rank in Google search results just by stuffing keywords into web content, so Google revised how its search engine processed content using numerous algorithms and NLP. NLP helps Google identify "spammy" content and categorize it. Google may de-index this content, penalize it, or simply rank it much lower than other content.

NLP is also used in email spam filters. Spammers try their best to evade such filters by changing words around, purposely misspelling words, or using synonyms. Email spam filters use a variety of factors to identify and block spam, phishing, and malicious content. Gmail's filter, for example, incorporates machine learning and NLP to perform "sentiment analysis." If content is determined to likely be spam, the content is sent to the user's junk folder. For some content, Gmail deletes the message.

A decade ago, application of NLP was comparatively complicated. AI-based technologies (including NLP and text analytics) have evolved considerably, and there are many cloud services, commercial products, and open source platforms businesses can leverage. Here are few open source NLP applications:

- Stanford CoreNLP
- Natural Language Toolkit
- Apache Lucene and Solr
- Apache OpenNLP
- GATE and Apache UIMA
- A Final Word

Text analytics isn't new, but it is still unfamiliar to many organizations. With APIs,cloud-based AI services, and open source platforms available today, your business can leverage the power of text analytics to get a competitive edge by better understanding your customers and improving your brand's value.

5. Flowcharts

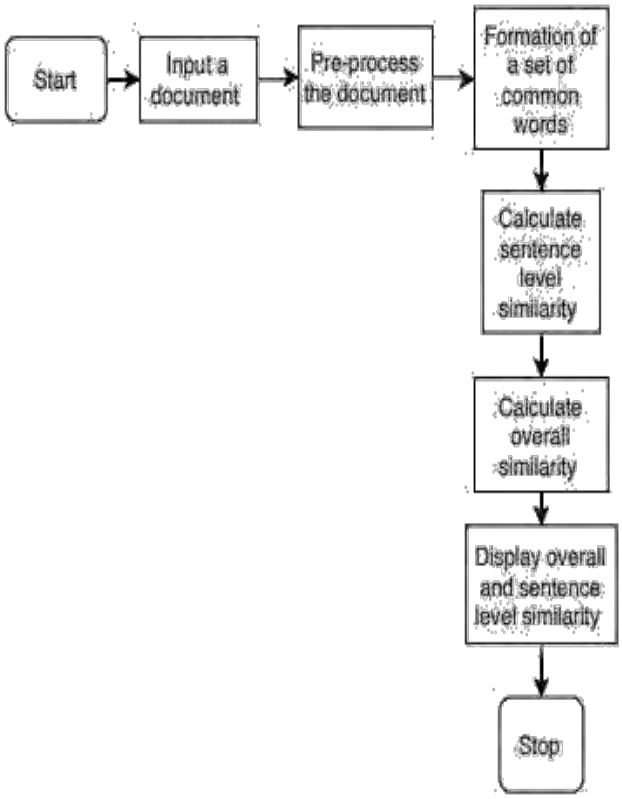


Fig 5.1 Syntactic Similarity

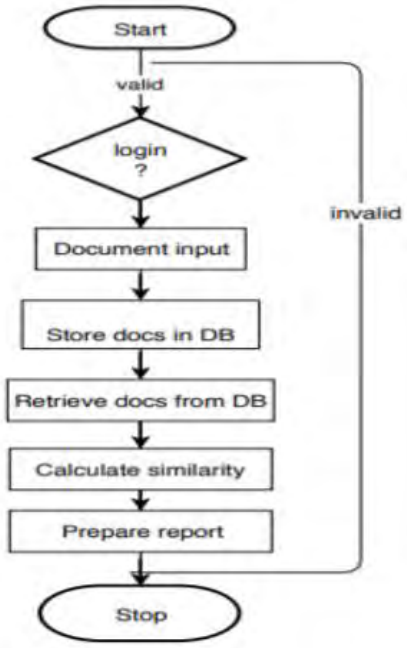


Fig 5.2 General flow diagram

V. RESULT AND DISCUSSION

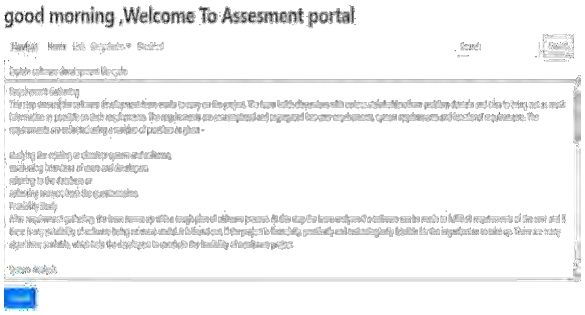


Fig 6.1 Teacher's Dashboard

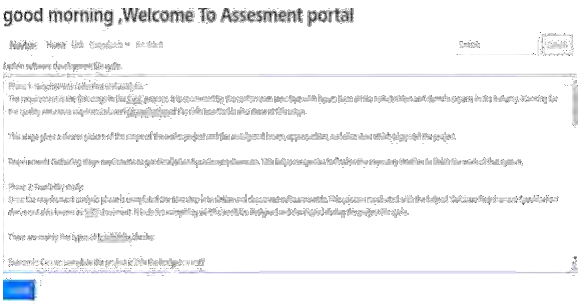


Fig 6.2 Student's Dashboard

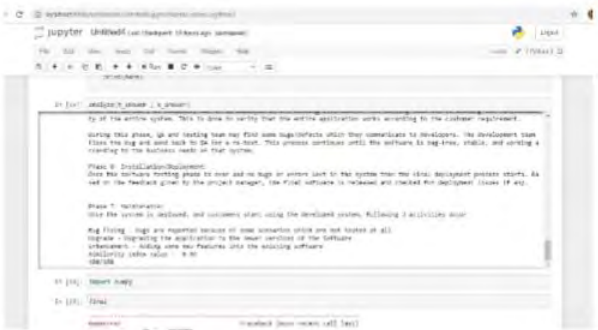


Fig 6.3 Result

VI. CONCLUSION

In this project we are using Natural Language Processing (NLP) technique for automatically evaluating the answer sheet in the form of text document for both objective and subjective type questions. The answer sheet is evaluated within the time limit. The mistakes that can occur during manual correction are reduced. Topics to be taught once again are also suggested. The work can be extended by parallelizing the evaluation process and reducing the time of evaluation.

VII. ACKNOWLEGEMENT

R.R. Sedamkar, Computer Engineering TCET, University of Mumbai, rr.sedamkar@thakureducation.org

Referance

[1] Piyush Patil, Sachin Patil, Vaibhav Miniyar, Amol Bandal, "Subjective Answer Evaluation Using Machine Learning" , Volume 118, No. 24, May 2018, International Journal of Pure and Applied Mathematics, Sinhgad Institute of Technology, Lonavala, Maharashtra,India.

[2] Bhagyashri Kaiche ,Samiksha Kalan, Sneha More, Lekha Shelukar4, " Online Descriptive Examination and Assessment System" , Volume 4, Issue 3, March 2014, International Journal of Emerging Technology and Advance Engineering, KBT College of Engg Nashik, Maharashtra, India.

[3] Nilima Sandip Gite , "Implementation of Descriptive Examination and Assessment System" , Volume No.7, Special Issue No.07, March 2018, International Journal of Advance Research in Science and Engineering, K. K. Wagh Polytechnic, Nashik, Maharashtra, India.

[4] Dharma Reddy Tetali, Dr. Kiran Kumar G, Lakshmi Ramana, PYTHON TOOL FOR EVALUATION OF SUBJECTIVE ANSWERS (APTESA), Volume 8, Issue 7, July 2017, pp. 247–255, Article ID: IJMET_08_07_029, International Journal of Mechanical Engineering and Technology (IJMET)

[5]Thomas N. T. Ashwini Kumar and Kamal Bijlani , Automatic Answer Assessment in LMS using Latent Symantic Analysis, Second International Symposium on Computer Vision and the Internet, Procedia Computer Science 58 (2015) 257 – 264 and Elsevier.

[6]Meena K and Lawrance Raj, Evaluation of the Descriptive type answers using Hyperspace Analog to Language and Self-organizing Map, 2014 IEEE International Conference on Computational Intelligence and Computing Research

[7]C. Leacock and M. Chodorow, "C-Rater: Automated Scoring of Short-Answer Question," Computers and the Humanities, vol. 37,no. 4, pp. 389-405, 2003.

[8] RaheelSiddiqi,Christopher J. Harrison and RosheenaSiddiqi,"Improving Teaching and Learning through AutomatedShort-Answer Marking", IEEE transactions on learning technologies, vol. 3, no. 3, JulySeptember 2010.

[9] SalvatoreValenti, Francesca Neri and Alessandro Cucchiarelli,"An Overview of Current Research on Automate Essay Grading"DIIGA Universita ' Politecnicadelle Marche, Ancona, Italy,Journal of Information Technology Education Volume 2, 2003.

Enhanced Interactive Intelligent Tutoring System

Dr. Anand Khandare, Rahul Yadav, Shubham Yadav, Mittal Jethva
Dept. of Computer Engineering
Thakur College of Engineering and Technology
Mumbai,India
parryrpy@gmail.com, ssubham.98@gmail.comm, mittaljethva.mj@gmail.com

Abstract— **Intelligent Tutoring System(ITS), is a computer-based self-learning system where students learn a specific topic. An Intelligent Tutoring System provides personalized learning content to students based on their requirements and preferences. And based upon their performance the analysis part will be done.**

ITS will contain a student's module and an adaptive engine. The student will be categorized based on bloom's taxonomy. The adaptive engine will use different pedagogy for each student to make the platform more interactive and challenging for the students. An adaptive system here, the next step of instruction which dependent upon student's performance. The objective of the system is to identify the capability of a student and the performance improvement needed by the student. And motivating the student for his/her learning process to give the appropriate encouragement, praise or reproach message. It is important to understand and represent the student's internal psychological states.

Keywords— ITS, Interactive, Motivation.

I. INTRODUCTION

Intelligent Tutoring System(ITS); is an important research area in which computer science and technology are used in education. It plays a very important role in Artificial Intelligence technology, in the absence of guidance from the human tutors to help learners acquire knowledge and different skills. Since from the '90s, many other countries have attached the great importance to intelligent tutoring systems research, and its application, and had made great achievements. However, with in-depth research, many problems came to light. For example, in the traditional ITS, the student is the passive recipient of teaching content. It lacks an interaction between systems and the students. Also, the traditional ITS model of the learners is not very successful, either too complicated to be handled, or toosimple that it lacks practical value. Thus, in the late 1990s, ITS researchers began to reconsider the application of artificial intelligence in ITS. Many new technologies have been applied to ITS, including multimedia, Internet-based distributed education, hypermedia technology. However, the introduction of new technology does not reduce the importance of artificial intelligence in teaching, on the contrary, these new technologies are of practical significance only when combined with artificial intelligence. This paper focuses on the application of artificial intelligence technology in ITS by describing the structure and function of the intelligent tutoring system and discusses several development trends of ITS.

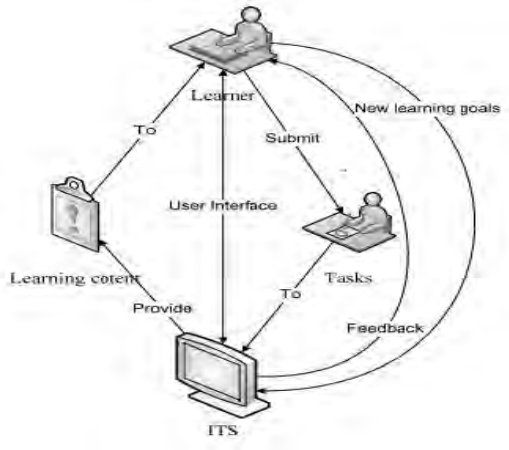


FIGURE 1. OVERVIEW OF SYSTEM FLOW

Technology-enhanced learning is a type of learning where self-motivation, communication, and efficiency are used. ITS is a system and technologies where adaptive learning technologies are used, which helps to individualize and personalize the learning process according to the individual's needs, analyze the knowledge of the students

Even a student's mood and emotions, as well as the learning style also get analyzed. Intelligent tutoring system work as a multi-agent system. Adaptive learning is the process of generating a unique learning experience for each student based on their personality, interests, and performance to achieve their goals such as academic improvement, learner satisfaction, effective learning process and many more. As the figure above shows, some important aspects of the implementation of the intelligent tutoring system to communicate with the student one's system need a strong user interface. This makes a student more comfortable while using the system .we can make the system more feasible and attractive to increase a student's comfort level.

II. LITERATURE REVIEW

Intelligent tutoring system(ITS) implementation can enhance the growth of students by the self-study method. There were much research has been done to implement these projects, And identify the limitations. Although it's difficult to implement the system by identifying each student's IQ level and also based upon that motivating student to perform more better, still

researcher is working on the problems. But this implementation will help a student to self-analyze and problem and work on the weak aspect by themselves. And by which student gets benefits to their academic as well as knowledge gaining skills.

Year	Author	Main Objective	Limitations
2019	Juhar Ahmed Abdella , Nazar Zakir[2]	An approach to identify the students' frustration in an ITS, inspired by the better accuracy of data-mining approaches, and the use of theory in modelling affective states.	Develop a based ITS and multiple basic questions make the scheme ITS.
2018	Craig, Scotty. [1]	ITSs typically operate by assessing a wide variety of learner characteristics and needs, sometimes at a very fine-grained level, and then adapting instruction according to principles of learning.	Difficult to and identify inner complexity.
.2013	Rajendran, Ramkumar & lyer, Sridhar & Murthy, Sahana & Wilson, Campbel & Sheard, Judy.[4]	An approach to identify the students frustration in an A limitation in the ITS, Inspired by the better accuracy of data-mining scalability of approach approaches, and the use of theory in modelling affective states.	A limitation scalability
2011	Praveen Kumar[4]	Develop web-based ITS and used multiple choice Instant feedback is absent questions to make the evaluation scheme independent of subject and studied different teaching-learning strategies.	In the document is quite small and in that project.
2009	Soldato, Teresa & Du Boulay, Benedict.	Discussion on motivation-based tactics and Not flexible contrast them with instruction .	Not flexible and the solution was not enough to rectify the problems.
2005	Yaratan Hüseyn.	Gives information about the existing four No solution provided for components of most IT expertise module, the problems and their some implementatio n strategies.	No provided for stated problems.

TABLE 1 TABULAR REPRESENTATION OF THE VARIOUS LITERATURE SURVEY

III. COMPONENTS OF SYSTEM

Like any other instructional system, the component represents the content to be taught, the device to understand and analyze the amount of knowledge that the student has, with the suitable teaching strategy, and a system for other communication issues. In computer-assisted instruction, all these components are in one structure whereas in ITSs they are separated and this gives more flexibility to the student and the machine in a way that they can have a one-to-one interaction just as when student and teacher sit down and attempt to teach and learn together (Park, Perez, and Seidel, 1986). Another advantage of these modules being separated shows itself when one attempts to make changes in the program; the change could be done only in one component of the program and the alteration of the whole programming content is not required. The three components mentioned below are referred to as

problem-solving or expertise module, student model and tutoring module, with the user interface.

A.EXPERTISE MODULE

An expertise module also called a problem-solving module consists of the domain knowledge that the system intends to teach to the student. The nature of knowledge is important while answering questions about intelligence. The Artificial intelligence community classifies knowledge into three categories i.e. conceptual, procedural, and imagine; and they worked with all three kinds of knowledge in the design of tutoring system programs. chaotic and nonlinear, problems have evolved out to be a better technique in capturing the structural relationship between the performance and its cognitive factors more accurately than many other statistical techniques. Basically in literature, different sets of input variables are used to predict value returns. Different input variables are used to predict the same set of returned data. Some researchers use i/p data from a single time series while others considered the inclusion of exclusive market information and macro-economic variables. Some researchers even pre-processed these input data sets before feeding it to any machine learning or AI algorithms.

B. STUDENT MODULE

The student model is used to assess the amount of knowledge that the student has in the material that is intended to be taught, to predict the learning behavior of the individual user and to identify the causes of errors. To evaluate the student model there are four major information sources which are:

- implicit (problem-solving behavior of the student),
- explicit (direct questions asked to the student),
- historical (assumptions based on the student's experience),
- structural (assumptions based on the difficulty of the subject-material).

This sophisticated procedure of creating a student model can be resolved to some extent by considering the student as a subset, simplification, or deviation of the expert's (computer's) knowledge. Simple pattern recognition is applied to a student's response history for making inferences about his understanding of the skill and his reasoning process used to derive the response.

C**. TUTORING MODULE**:

Once the student model has been selected and the expertise in the subject domain is given, the expert tutor selects an initial efficient teaching strategy based on the student's previous performance. This teaching strategy is modified as the student model evolves and the pedagogical strategies used may include presenting increasingly complex concepts or problems. The fundamental issues for a tutor as listed by Dede are:

Whether to intervene in the information flow,
What to discuss,
Which presentation strategy to use,
How much to say.

IV. DATA COLLECTION

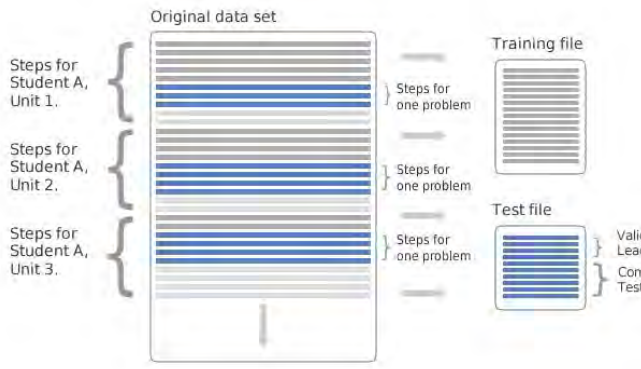


FIGURE 2 ORGANIZATION OF DATASET

Attribute	***Description of Attributes***
Subject ID	Id number of a subject.
Student ID	Id number of student.
Subject Name	Name of a subject chosen.
Student Name	Name of a student.
Last Attempted	Last question which student has attempted.
Number Of Attempts	Total number of questions which being answered.
Course Completion time	Time taken by the student to complete a particular course.
Average Time	Average time required by all students .
Hint For reference	Hint provided for the student for reference.
Level of student	In which level student has been performing.
Solution Of Questions	Solutions provided for each questions in the topic.
Confidence Test Score	Total scored obtained in confidence test which is basically full length test.
Attempts in Confidence Test	Total number of attempted question in confidence test.

Average Score of Confidence Test	Total average score of confidence test of all students.
Motivational Data	Slogan's or quotes which help the students to boost their confidence.

TABLE 2
DESCRIPTION OF ATTRIBUTES INVOLVED IN THE DATASET

V. ALGORITHM USED IN SYSTEM

A. Case-Based Reasoning

Case-based reasoning, or training examples, where the cases, are stored and accessed to solve new problems. To get a prediction for nearly new example, those cases that are similar, or very close to the new example are used to predict the targeted value of the new example. This is one of the extremes of learning problems where, unlike the decision trees algorithm and neural networks, relatively little work must be done offline, and virtually all of the work is performed at execution time. Case-based reasoning is used for classification and regression techniques for data. It is also applicable when the case is complicated, such as in legal cases, where the cases are complex legal rulings, where the cases are difficult to give solutions to complex problems.
For this method to work, a distance metric is required to measures the distance between two close examples. First define a metric for the domain of each feature, in which the values of the features are converted to a numerical scale that is used to compare values.

$$d(e_1, e_2) = \sqrt{\sum_i w_i * (X_i(e_1) - X_i(e_2))^2}.$$

The weights could be provided as an input. It is also possible to learn the weights. The learning agent would try to find weights that minimize the error in predicting the value of each component of the training set, based on every other instance in the given training set.

B. Support Vactor Machine

The Support Vector Machine (SVM) is a supervised machine learning algorithm that is used for both classification or regression problems. However, it is mostly used in classification problems. In this algorithm, we plot the data item in on an n-dimensional space with the value of each feature being the value of a particular co-ordinate. Then, we perform a classification operation by finding the hyper-plane that differentiates the two classes. Support Vector Machine is primarily a classifying method that performs classification tasks by creating multidimensional spaces that separate the different cases of different class values.
This system supports both regression and classification tasks and it can handle multiple continuous and categorical values. For categorical variables, a dummy variable is well created with case values as either 0 or 1. For this type of SVM, training are involves the minimization of the error function:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

Subject to the constraints

$$y_i \left(w^T \phi(x_i) + b \right) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \ldots, N$$

Here C is the capacity constant, w is the vector of a coefficient, b is a constant, and represents parameters for handling no separable or input data. The index i labels the different N training cases. Here data represents the class labels and xi represents the independent variables. The kernel is generally used to transform data from the input to the feature spaces. It has been noticed that the as much larger the C, that much more the error is detecting. Thus, C should be select with care to avoid overfitting data. In contrast to Classification support vector machine Type 1, the Classification Type minimizes the error function:

$$\frac{1}{2} w^T w - v\rho + \frac{1}{N} \sum_{i=1}^{N} \xi_i$$

subject to constraints forms:

$$y_i \left(w^T \phi(x_i) + b \right) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, \ldots, N \text{ and } \rho \geq 0$$

In a regression support vector machine, you have to estimate the functional dependence of variable y on a set of the independent variables x. It assumes that, like other regression problem, that the relation between independent and dependent variables is given by a function f plus the addition of some additive noise values.

VI. CONCLUSION

We analyzed the requirements and objectives of our project. We created a basic block diagram representing the functionality of our project also, we analyzed software to be used for efficient execution. We came across various algorithms and datasets which will be beneficial for our project. We are currently working on identifying more algorithms suitable for our project. We are also working on dividing our project into modules and working collectively on each module. These modules will be the basic building blocks of our project which will be coded and integrated later.
This project is only the first step towards achieving that goal. This modular architecture ensures that every module can be extended and developed separately. It opens up a lot of different approaches for future research and development purposes.

VII. ACKNOWLEDGMENT

We sincerely thank our guide Dr. Anand Khandare for his guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

References

[1] Rajendran, Ramkumar & Iyer, Sridhar & Murthy, Sahana & Wilson, Campbell & Sheard, Judy. (2018). *A Theory-Driven Approach to Predict Frustration in an ITS. Learning Technologies*, IEEE Transactions on. 6. 378-388. 10.1109/TLT.2018.31.

[2] Soldato, Teresa & Du Boulay, Benedict. (1996). *Implementation of Motivational Tactics in Tutoring Systems. Journal of Artificial Intelligence in Education*. 6. 337-378. 10.1007/s40593-015-0052-1.

[3] Dede, C. A Review and Synthesis of Recent Research in Intelligent Computer-Assisted Instruction. Int. J. ManMachine Studies 24, 329-353. Academic Press Inc. (London) Limited,1986.

[4] Klein, M.F. Curriculum Design. International Encyclopedia of Education: Curriculum Studies. Volume II. Husen, T. and Postlethwaite editors, Pergamon Press: Oxford England, 1985, pp1163-1170.

[5] Bloom, B.S., The 2 Sigma Problem: The Research for Methods of Group Instruction as Affective as One-to-one Tutoring. Educational Researcher, 13, 3-16. 1984.

[6] Bloom, B.S., et. al. Taxonomy of Educational Objectives: Cognitive Domain. David McKay Co.: New York, 1956.

[7] Carr, B., and I.P. Goldstein. Overlays: A Theory or Modeling for Computer-Assisted Instruction (MIT AI Lab Memo No. 406). Cambridge, MA: MIT, 1977.

[8] Dede, C. A Review and Synthesis of Recent Research in Intelligent Computer-Assisted Instruction. Int. J. ManMachine Studies 24, 329-353. Academic Press Inc. (London) Limited, 1986.

Implementation Of Eye Trackingbased InteractionSystem Fordisabled Users

Tushar P.Rajdev, AjayV.Mule, Sakshi S.Pandey, Dr.Sheetal Rathi
Computer Department Thakur College of Engineering and Technology, Mumbai,India
tushar.rajdev@gmail.com, muleajay08@gmail.com, pandeysakshi388@gmail.com, sheetal.rathi@thakureducation.org

Abstract—**Eye-trackinghasbeen widelyused inmedicaland psychological research, forobserving and collectingeye behaviorsuch aswhendo we blink,pupildilationand movement. Concerningcomputervision,itisprimarilyused to measure the pointof gaze,i.e.where the user islooking.Real-timeeyetrackingprovideseffectivemeansofinteraction and communicationfor peoplewithphysical disabilities. Inthis paper,weproposeasystemwhichfocusesonprovidingasimple and convenientmethod for easyinteraction withthesystem usingusers'eye(pupil).It canbeused to controlacomputer with youreyesinstead ofusinga traditionalkeyboard and mouse,enablingindividualswithphysicaland cognitive disabilitiestolivericher andmoreindependentlives. Thereare someexistingproductsthatprovide eye tracking, butthose usuallyrequiresusinghigh endvideo recordingdevices.So in anattempttoprovideasystemthatcantrackeyeswiththebasic requirements ofapersonalcomputer or laptopwithaweb cam (withinfraredlight,ifneeded),we have writtenthis paper.**

Keywords—***Eye tracking,Gazetracking, Eye pupil detection, Human ComputerInteraction, Eye movement,Video based eye movement***

I. INTRODUCTION

Intoday'sworld,computers havebecomean integral partof ourdaily life, weuseitonadaily basis.But, manyarephysically disabledandhencearen'tcapableof usingthistechnology.Theusageofcomputersbythemwould notonlyhelpthemcommunicateand connectwiththeworld butcanalsoeasetheirlife.Thisprojectaimsatcreating a systemthatusesEyemovementtracking tocontrol the computersystem.

Eye-trackingreferstotheprocessofestimatingwherewe look, alsoknownasourpoint ofgaze.Thesemeasurements arecarried outbyaneyetracker,whichrecords theposition oftheeyesandthe movementstheymake.

Near-infrared lightiscoordinatedtowardsthecenterof theeyes(pupil),causingdistinguishable reflectionsinboth thepupilandthecornea(theouter-mostopticalelementofthe eye).Thesereflections –thevector betweenthecorneaand thepupil–aretracked byaninfraredcamera. Thisisthe opticaltrackingofcornealreflections,knownas pupil center cornealreflection(PCCR)

Aninfraredlightsourceisrecommended astheprecisionof gazedirectionestimation issubjecttoacleardemarcation (anddetection)ofthepupilaswellasthedetectionofcorneal reflection.Normallightsourcesaren'tcapabletogiveas Muchcontrast,implyingthatasuitableamountofaccuracyis muchdifficulttoachievewithoutinfraredlight.Lightfromth evisiblespectrum isprobably goingto generateuncontrolledspecularreflection,whileinfraredlig ht takesintoconsideration aprecisesseparationbetweenthe pupilandtheiris–whilethelightdirectlyentersthepupil,it just"ricochets off"theiris.Moreover,asinfraredlightisn't visibletohumansitdoesnotcauseanyinterruptionwhilethe eyesarebeingtracked.Acommonmethodofcalculating thefocusoftheuser's gazecomparesthepositionofanear-infraredlight (reflected intheeye)totheposition ofthepupil.Thisinformation, combinedwiththeposition oftheparticipant's head,canbe extrapolatedtodeterminethepointtheuser'seyesarefocuse d on,andtherebythecorrespondingscreencoordinates.

Whiletherearevarious eyetrackers, therearetwomain types: screen-basedand glasses.They areusedacross avariety offieldsandresearchareas,howeverhowtheyare utilizedandthedatatheyprovidesubsequentlycandiffer.

II. LITERATURESURVEY

Faceistheindexofmind andeyesarethewindowtothe soul.Eyemovementsgivearichandinformativewindowinto anindividual'sthoughtandintentions.Thusthestudyofeye movementmayfigureoutwhat people arethinkingbasedon wheretheyarelooking.Eye-trackingistheestimationofeye movementandgazetracking istheanalysis ofeye-tracking informationwithrespecttothevisualscreen.Researches of thisfieldregularly usethetermsgaze-tracking,eye-tracking oreye-gazetrackingconversely.Eye-trackingismostlyused inapplicationslikediagnosis ofvariousclinicalconditions, drowsinessdetectionoreveniris recognition.Howevergaze trackingmethodscanbeused inallthewaysthatoneutilizes theireyes,to nameafew,eyetypingforphysicallydisabled.

Generally,theintegration ofeyeandheadpositionis utilized tocompute theareaofthegazeinthevisualscreen. Simpleeyetrackersreportjustthedirection ofthegaze relativetotheheadorforafixedpositionoftheeyeball.Such eye-tracking systemsarereferredtoasinvasive orintrusive systemsbecausesomeuniquecontacting gadgets are connected totheskinoreyetogettheuser'sgaze.The systemswhichdonothaveanyphysicalcontactwiththeuser andtheeyetrackercontraption arereferredtoasremote systemsornon-intrusivesystems.

The most broadly used current designs are video-based
Eye-trackers. A camera focuses around one or both eyes and records eye movement as the individual gazes at some kind of stimulus. Most modern eye-trackers use the focal point of the pupil and near-infrared non-collimated light to make corneal reflections (CR). The vector between the pupil centre and the corneal reflections can be utilized to calculate the point of gaze on the surface or the gaze direction. A basic calibration process of the individual is usually required before using the eye tracker.

Two general types of infrared / near-infrared eye-tracking methods are used: bright-pupil and dark-pupil. Their distinction depends on the location of the illumination source with respect to the optics. In the event that the illumination is coaxial with the optical path, then the eye acts as a retro reflector as the light reflects off the retina making a bright pupil impact similar to red-eye. If the illumination source is counter balanced from the optical path, then the pupil appears dark because the retro reflection from the retina is coordinated away from the camera. The Eye gaze System must be used in an environment where there is less infrared light. The System works best away from windows, and in a room lit with fluorescent or mercury-vapour lights, which are low in infrared.

In this research survey, various algorithms, models and methods have been studied. Based on review, certain significant parameters and components that affect the precision and efficiency of eye gaze techniques have been recognized. Different eye gaze-based HCI models and algorithms have been read for additional examination and investigation. Early work on intuitive eye-tracking applications depends primarily on users with inabilities but can also be used for general users managing with portable computers.

III. PROPOSED WORK

Eye gaze or eye tracking is a technique of accessing your computer or communication aid using a mouse, controlled with your eyes. The systems follow your eyes with astonishing precision to see where you are looking on the screen. You would then be able to choose the item you are looking at by dwelling, blinking or clicking with a switch.

The project works by having lights and cameras that are continually sending and receiving data. The camera gets light reflections from your pupils and interprets the movement of your eyes into mouse cursor movements. It takes just seconds to finish a one-time calibration.

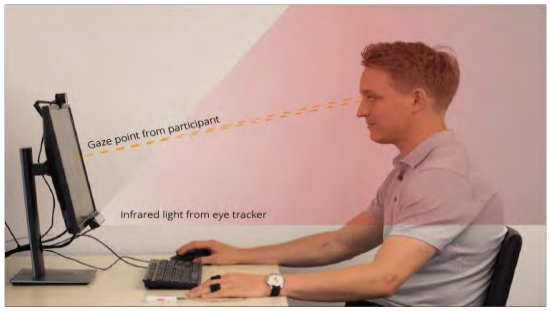


Figure 1: Proposed working Mechanism

As a user sits in before the screen, a specialized video camera mounted above/below the screen observes one of the

User's eyes. Advanced image-processing algorithm in the system consistently examines the video image of the eye and figure out where the user is looking on the screen. Nothing is appended to the user's head or body. In detail the process can be described as follows:

The Project uses the corneal reflection/pupil-centre method to figure out where the user is looking on the screen. As infrared-sensitive video camera, mounted underneath the System's screen, takes 60 pictures for each second of the user's eye. A low power, infrared light-emitting diode (LED), mounted in the centre of the camera's focal point illuminates the eye. The LED reflects a little bit of light off the surface of the eye's cornea. The light also radiates through the pupil and reflects off of the retina, the back surface of the eye, and causes the pupil to appear white. The bright-pupil impact improves the camera's image of the pupil and makes it simpler for the image processing functions to find the centre of the pupil. The computer calculates the person's gaze point, i.e., the direction of where he is looking on the screen, on the basis of the relative positions of the corneal reflection and pupil centre within the video image of the eye.

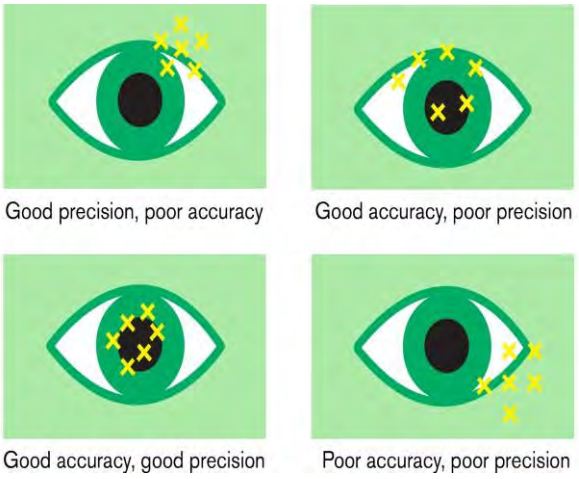


Figure 2: Eye precision and accuracy

500mm	7.4mm (0.70)	7.9mm (0.75)	10.8mm (1.0)
520mm	12.5mm (1.15)	16.2mm (1.45)	20.4mm (1.9)
580mm	12.1mm (1.03)	15.1mm (1.29)	19.3mm (1.7)

Table2:Accuracy Table

Typicallythesystempredictsthegazepointwithanormal precisionofaquarter-inch orbetter.Beforeoperating,the eye-tracking applications,thesystemmustlearnafew physiological propertiesofauser'seyesoastobeableto displayhisgaze-point precisely.TheSystemlearnsthese propertiesbyperforminga calibrationprocess.The user calibratesthesystem byfixinghisgazeonsmallcircles displayed onthescreenandfollowingitasitmovesaround thescreen.Thecalibrationprocessgenerally takesabout15 seconds,andtheuserneednotrecalibrate ifhemovesaway fromthe systemandreturnslater.

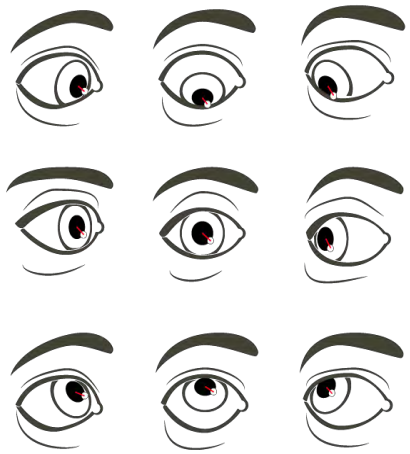


Figure3:PupilCentre cornealMovement

A. Applications

Controllingofmousecursorandkeyboardinputs withonlyeye movements
Surfthewebandsendemails

Operatelightsandappliancremotely(Byaddin g homeautomation)

Control otherwirelessdevicesconnectedtothe computer

Creating afastandefficientwayforusingthe computer

B. Methodology

Initially,faceandeyedetectionisdonetostart.Wedothis inordertoreducetheprocessingtimeandfocusmainlyonthe requiredfeatures offaceandeye.Inorderdoobject recognition/detection withcascadefiles,youfirstneed cascadefiles.Fortheextremely populartasks,thesealready exist.Detecting thingslikefaces,cars,smiles,eyes,and licenseplates,forexample,areallprettyprevalent.Inthispro ject,weappliedfacedetectiontosomephotos

Distanceto the Camera	Horizonta accuracy (Measure inDegree)	Vertical Accuracy (Measurein Degree)	Total Accuracy (Measurei Degree)
360mm	7.4mm (0.91)	9.7mm (1.21)	12.2mm (1.5)
380mm	4.3mm (0.51)	6.1mm (0.73)	7.4mm (0.89)
410mm	4.1mm (0.46)	7.2mm (0.80)	8.2mm (0.93)
440mm	4.2mm (0.46)	6.7mm (0.71)	7.9mm (0.84)
470mm	8.8mm (0.88)	6.6mm (0.67)	11mm(1.1)

wetookusingOpenCV withPython. OpenCVisanopen-sourcesoftwarelibrary thatallowsdevelopers toaccess routinesinAPI(ApplicationProgrammingInterface)usedf or

Computer visionapplications. The versionweusedwasdevelopedforPythoncalledOpens-Python Eye-trackingrefers totheprocessof estimatingwherean individual looks,alsoknownastheirpoint ofgaze.These measurementsarecompletedbyaneyetracker,whichsaves thepositionoftheeyesandthe movementsthey make.Near-infraredlightiscoordinated towardsthecenteroftheeyes, causingperceptiblereflectionsin boththe pupiland the cornea. Thesereflections –thevectorbetweenthepupiland thecornea–aretracked byaninfraredcamera. Thisisthe optical tracking ofcorneal reflections,alsoknownaspupil centercornealreflection.

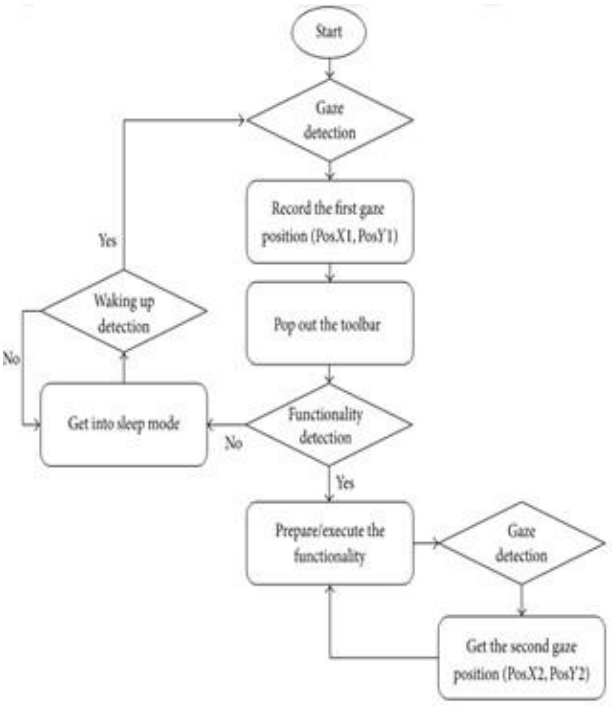


Figure4:Workflowofthe system

Distanceto the Camera	Accuracywith basicwebcam	Accuracywith infraredcamera
360mm	12.2mm(1.45)	8.4mm(0.96)
380mm	7.4mm(0.82)	5.2mm(0.63)
410mm	8.2mm(0.90)	6.0mm(0.68)
440mm	7.9mm(0.86)	5.5mm(0.61)
470mm	11mm(1.31)	7.7mm(0.85)
500mm	10.8mm(1.30)	6.9mm(0.76)

Aninfraredlightsourceisrecommendedas the precision ofgazedirectionmeasurement isreliantonaclear demarcation ofthepupilaswellasthedetectionofcorneal reflection.Normallightsourcesarenotcapableof providing asmuchcontrast,implyingthatasubstantial amountof accuracyismuchdifficulttoachievewithoutinfraredlight.Li ghtfromthevisiblespectrum islikelytocreate uncontrolled specular reflection,whileinfraredlightallows forapreciseseparationbetweenthepupilandtheiris–while thelightdirectlyentersthepupil,itjust"ricochets off"the iris.Additionally, asinfraredlightisn'tvisibletohumansit doesnotcauseanyinterruption whiletheeyesarebeing tracked.

Characteristicsofproposedsystem:
Existing systemrequireshigh featuredhardwaresystem,proposedsystemwould provideextremely lightweightdevicesthatarereliableandconvenient forusertouse.
The proposedsystemisorientedtowards the possibility ofbeingusedwidely,whichsupports mostofthelow-costeyetrackersthatareaffordable forthe majorityofusers.

Theproposedsystemfulfilsallofthefunctionsof regularinputsources(Mouse,VirtualKeyboard etc.), and the user will be able to interact with the system efficiently.

 the proposed system provides more natural and more convenient communication mechanism for user-computer dialogue and could also avoid annoying user with unwanted responses to their actions.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Although not in a full working condition, the proposed methodology of our project provides a good accuracy, with a stable precision of the gaze point. Experimented in various light conditions, the system works more efficiently in scenarios where there is less illumination from behind the user. Our goal to create a light weight and low cost system, with only the basic requirement of personal computer or laptop with a web cam, to an extent has been successful, with an additional requirement of Near Infrared (NIR) camera, the accuracy and precision of the system increases vastly.
The comparison is shown in the table below:

From the above table it is clearly evident that using an infrared based camera provides better accuracy than just a basic webcam. Also the difference in precision of the gaze point between the two methods is vast. Hence, we have decided on using a NIR camera, to get an almost accurate result
Snapshot of sample working is shown below:

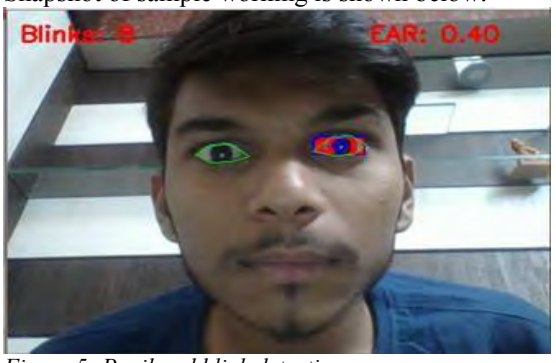


Figure 5: Pupil and blink detection

Figure 6: Thresholding and contour detection

ACKNOWLEDGEMENT
We wish to express sincere thanks to Xuebai Zhang et al for their research paper on *Eye Tracking Based Control System for Natural Human-Computer Interaction 2017* for providing a backbone for our project. We also extend our heartfelt thanks to Adrian Rosebrock for his website *PyImageSearch* which has helped us to understand the basics and even some expert level concepts of computer vision.

References

[1] Xuebai Zhang, Xiaolong Liu, Shyan-Ming Yuan, and Shu-Fan Lin, "Eye Tracking Based Control System for Natural Human-Computer Interaction", *Department of Computer Science, National Chiao Tung University, Hsinchu 300*
[2] Kyung-Nam Kim, R.S. Ramakrishna, "Vision Base Eye Gaze Tracking For Human Computer Interaction*", Department Of Information And Communication, Kwangju Institute Of Science And Technology, Kwangju, 300-712, Korea (ROK).*
[3] Singh, Hari. "Human Eye Tracking and Related Issues: A Review". *International Journal of Scientific and Research Publications. 2. 1-9.*
[4] Anna Maria Feit et al, "Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design", *Aalto University & Microsoft Research*
[5] Manu Kumar, "Gaze-Enhanced User Interface Design, A Dissertation Submitted To The Department of Computer Science And The Committee on Graduate Studies*, Stanford University, 2007.*
[6] Linda E. Smibert And James N. Temple man, Robert J. K. Jacob, "Evaluation and Analysis of Eye Gaze Interaction", *Human Computer Interaction Laboratory, Naval Research Laboratory*
[7] Washington, Dc 20375, Department Of Electrical Engineering and Computer Science, *Tufts University Medford, Ma 02155.*
[8] Jowers Iestyn, Prats Miquel, Mckay Alison, Steve Garner, "Evaluating An Eye Tracking Interface For A Two Dimensional Sketch Editor", *University of Leeds, Leeds And The Open University, Walton Hall, Milton Keynes, Mk7 6aa, Ls2 9jt, UK.*
[9] Susanna Nilsson, Torbjorn Gustafson and Per Carlsberg, "Hands Free Interaction with Virtual Information in A Real Environment:

Implementation of a System for Winner Predictor Using Machine Learning

Ayush Agarwal, Bhavya Agrawal, Nitin Harane, Dhruv Bangur
Department of Computer Engineering
Thakur College of Engineering and Technology Mumbai, India
ayushgagarwal@gmail.com, agrawal.bhavya8@gmail.com, nitinharane@thakureducation.org, bangurdhruv@gmail.com

Abstract—**Cricket, particularly the Twenty20 configuration, has greatest vulnerability, where a solitary over can totally change the force of the game. With a great many individuals following the Indian Premier League (IPL), building up a model for anticipating the result of its matches is a genuine issue. A cricket coordinate relies on different components, and right now, factors which fundamentally impact the result of a Twenty20 cricket coordinate are distinguished. Every player execution in the field is considered to discover the general weight (relative quality) of the group. A multivariate relapse based arrangement is proposed to ascertain purposes of every player in the class and the general load of a group is figured dependent on the past exhibition of the players who have showed up most for the group.**

Right now, recent years information of IPL containing the players subtleties, coordinate scene subtleties, groups, ball to ball subtleties, is taken and broke down to make different inferences which help in the improvement of a players execution. Different highlights like how the scene or hurl choice has impacted the triumphant of the match in most recent ten years are additionally anticipated. AI and Data extraction models are considered for forecast is Random Forest Regressor. AI models were prepared and utilized for anticipating the result of each 2018 IPL coordinate. We will perform three types of prediction toss-winner prediction, 2nd inning score prediction and the actual winner prediction. The forecast outcomes are great.

Keywords— Regression, Random Forest Regressor, Prediction, Data Extraction

I. INTRODUCTION

Machine learning is a branch of artificial intelligence that targets at resolving real life engineering issues. It gives the chance to learn without being unequivocally customized and it depends on the idea of gaining from information. It is so much predominantly used dozen a times a day that we may not even know. The upside of machine learning (ML) techniques is that it uses mathematical models, heuristic learning, knowledge acquisitions and decision trees for decision making. In this manner it provides controllability, observability and stability.

The sport of cricket is played in different configurations, i.e., One Day International, T20 and Test Matches. The Indian Premier League (IPL) is a Twenty-20 cricket tournament league built up with the goal of advancing cricket in India and in this way supporting youthful and skilled players. The league is an annual event where teams representing various Indian cities go up against one another. The teams for IPL are selected by means of an auction. Players' auctions are not a new phenomenon in the sports world. However, in India, selection of a team from a pool of available players by means of auctioning of players was done in Indian Premier League (IPL) for the first time. Due to the involvement of money, team spirit, city loyalty and a massive fan following, the outcome of matches is very important for all stake holders. This, in turn, is dependent on the complex rules governing the game, luck of the team (Toss), the ability of players and their performances on a given day. Various other natural parameters, such as the historical data related to players, play an integral role in predicting the outcome of a cricket match. A way of predicting the outcome of matches between various teams can aid in the team selection process . However, the varied parameters involved present significant challenges in predicting accurate results of a game. Besides, the accuracy of a prediction depends on the size of data utilized for the same.

Further, several predictive models are also built for predicting the result of a match, in light of every player's past performance just as some match related data. The developed models can help decision makers during the IPL matches to access the strength of a team against another. The contributions of the presented work are as per following:

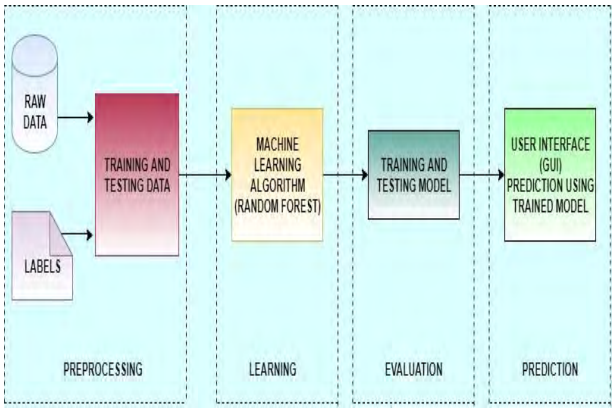
To predict the toss winning team of the match based on the city in which the match is played.

To predict the performance of a team in the 2 inning depending on the run made, wickets fallen and numbers of overs left.

To successfully predict the outcome of IPL matches.

II. METHODOLOGY

The overall working of our project can be explained through the following abstract diagram.



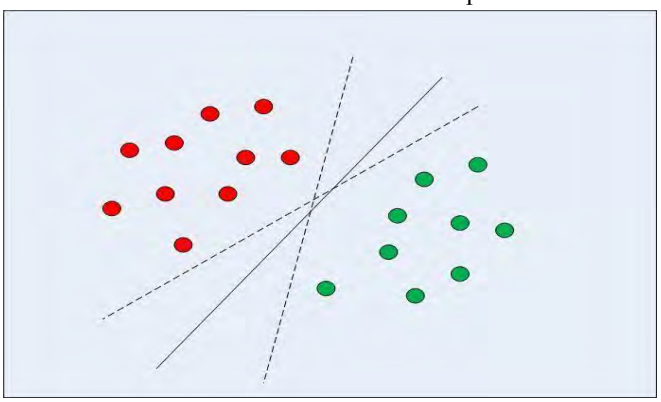
The methodology consists of 3 main stages- Data Pre-processing, Data Preparation, Encoding the data. Initially, the IPL dataset of 10 years is taken. In data pre-processing phase, the data is incomplete, noisy and inconsistent. The data preparation is significant for achieving optimal results. This involves choosing an outcome measure to evaluate different predictor variables. In data Encoding phase, label each term with short names and encode them as numerical values for predictive modelling as implemented below. matches.replace (['MumbaiIndians','KolkataKnightRiders','RoyalChallengersBangalore','DeccanChargers','ChennaiSuperKings','RajasthanRoyals','DelhiDaredevils','GujaratLions','KingsXIPunjab','SunrisersHyderabad','RisingPuneSupergiant','KochiTuskersKerala','PuneWarriors'],['MI','KKR','RCB','DC','CSK','RR','DD','GL','KXIP','SRH','RPS','KTK',PW'],inplace=True)encode={'team1':{'MI':1,'KKR':2,'RCB':3,'DC':4,'CSK':5,'RR':6,'DD':7,'GL':8,'KXIP':9,'SRH':10,'RPS':11,'KTK':12,'PW':13},'team2':{'MI':1,'KKR':2,'RCB':3,'DC':4,'CSK':5,'RR':6,'DD':7,'GL':8,'KXIP':9,'SRH':10,'RPS':11,'KTK':12,'PW':13},'toss_winner':{'MI':1,'KKR':2,'RCB':3,'DC':4,'CSK':5,'RR':6,'DD':7,'GL':8,'KXIP':9,'SRH':10,'RPS':11,'KTK':12,'PW':13},'winner':{'MI':1,'KKR':2,'RCB':3,'DC':4,'CSK':5,'RR':6,'DD':7,'GL':8,'KXIP':9,'SRH':10,'RPS':11,'KTK':12,'PW':13,'Draw': 14}}
The test dataset is taken as input to the learning algorithm, evaluates different scenarios like toss winner, toss decision and team winner. Thus, new data is obtained with final prediction.

A. Support Vector Machine

A support vector machine is a type of supervised machine learning classification algorithm. SVMs were first seen in
1960s and were later refined in 1990s. However, they are becoming somewhat popular because of their ability to achieve brilliant results. SVMs are implemented in a different way when we compare them to other machine learning algorithms.

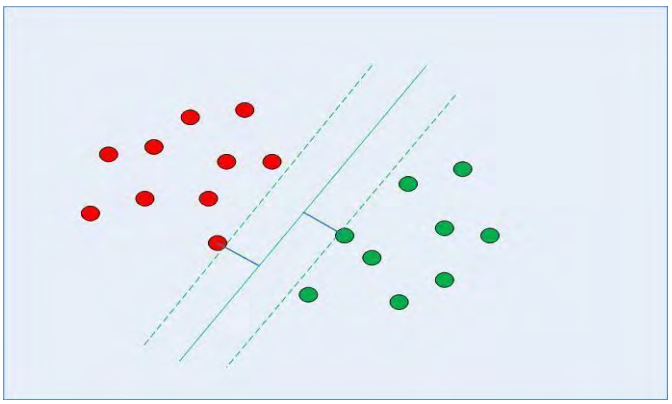
In case of linearly separable data in two dimensions, as shown, a typical machine learning algorithm tries to find a boundary that divides the data in such a way that the misclassification error can be minimized. There can be several boundaries that correctly divide the data points. The presence of dashed lines and solid line is used to classify the data.

SVM differs from the other classification algorithms in the way that it chooses the decision boundary that maximizes the distance from the nearest data points of



all the classes. An SVM finds the most optimal decision boundary that can be made.

The most optimal decision boundary is the one which has maximum margin from the nearest points of all the classes. The nearest points from the decision boundary that maximize the distance between the decision boundary and the points are called support vectors as seen in the figure. The decision boundary in case of support vector machines is called the maximum margin classifier, or the maximum margin hyper plane.



We identified certain problems with SVM which is why we could not use it in the project. Some of them have been mentioned below.

1. SVMs do not perform well on highly skewed/imbalanced data sets. These are training data sets in which the number of samples that fall in one of the classes far outnumber those that are a member of the other class. Customer churn data sets are typically in this group because when you collect the training set, among a million customers during a particular time period, there would be very few who have actually churned.

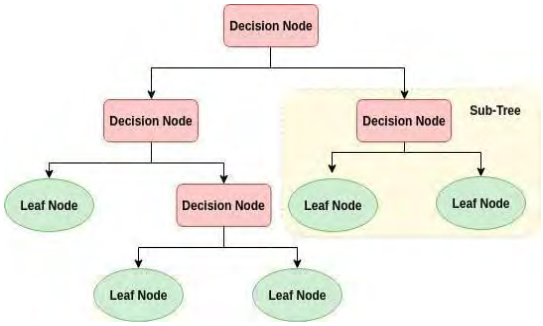
2. SVMs will not work well in multiple classes. They should not be used when there are a number of classes. Then you have to get back to a binary classifier itself and a voting mechanism to classify the classes among all the classes that are at hand.

3. The number of training samples should be much greater than the number of features. SVMs will not

work otherwise. Therefore, we need a lot of training samples if we want to use SVM

B. Decision Tree Classifier

A decision tree is a tree structure which is like a flowchart, where an internal node represents an attribute or a feature, the branch represents a decision rule, and outcome is represented by each leaf node. Root node is the topmost node of the tree. Partition can be made on the basis of attribution value. It partitions the tree in recursively manner call recursive partitioning. Decision making is the use of this flowchart like structure. It's visualization like a flowchart diagram which easily mimics the human level thinking. Decision trees are easy to understand and interpret, that is why they can be used too.



White-box models are the type of models which one can clearly explain how they behave, how they produce predictions and what the influencing variables are. White box has two properties or features that they have to be understandable and transparent as well. Decision tree is white box example. Unlike Neural Network and black box, it shares internal decision making logic. Neural Network algo is slower training time as compared to this. The time complexity of decision trees is a function of the number of records and number of attributes in the given data. Distribution-free or non-parametric method is used in decision tree, which does not depend upon probability distribution assumptions. High dimensional data with good accuracy is an important feature of decision tree.

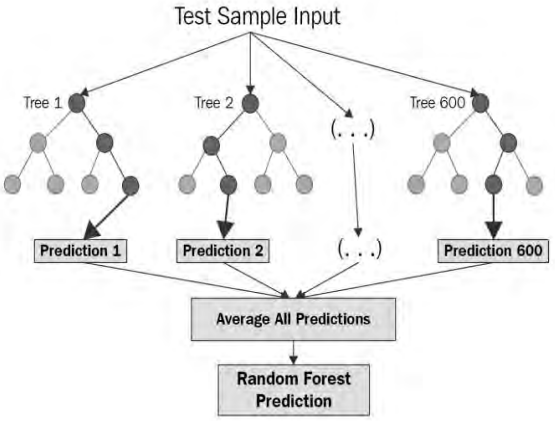
We identified certain problems with SVM which is why we could not use it in the project. Some of them have been mentioned below.

1. Small changes in data brings large change in structure causing instability.
2. For a Decision tree sometimes calculation can go far more complex compared to other algorithms.
3. Decision tree often involves higher time to train the model.
4. Decision tree training is relatively expensive as complexity and time taken is more.

C. Random Forest

Random forests is a supervised learning algorithm. Classification and regression can be done using random forest. Most flexible and easy to use algorithm is random forest. Forest is made of many trees. Forest becomes more robust with an increase in number of trees. Decision trees are made on randomly selected data samples, then it get a prediction form from each tree, and the best solution bymeans of voting. Feature importance factor is also provided by the random forests.

Having a variety of applications, such as recommendation engines, image classification and feature selection, it can be used to classify loyal loan applicants, identify fraudulent activity and predict diseases. Accuracy and robustness in provided by the random forests because of the number of decision trees participating in the process. Overfitting problem is avoided in the random forests too. Biases are cancelled out very nicely due to the average and hence it increases the accuracy of the prediction. The algorithm can be used in both classification and regression problems. Random forests can also handle missing values. There are two ways to handle these: using median values to replace continuous variables, and computing the proximity-weighted average of missing values. You can get the relative feature importance, which helps in selecting the most contributing features for the classifier.



Random forest is a bagging technique but not a boosting technique. The trees in random forest run in parallel sequence. There is no interaction among the trees while building them. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

A random forest is a meta-estimator (i.e. it combines the result of multiple predictions) which aggregates numerous decision trees, with some supportive adjustments:

1. The number of features that can be split on at each node is limited to some percentage of the total (which is known as the hyperparameter). This guarantees that the ensemble model does not depend too intensely on any individual

feature, and makes fair use of all potentially predictive features.

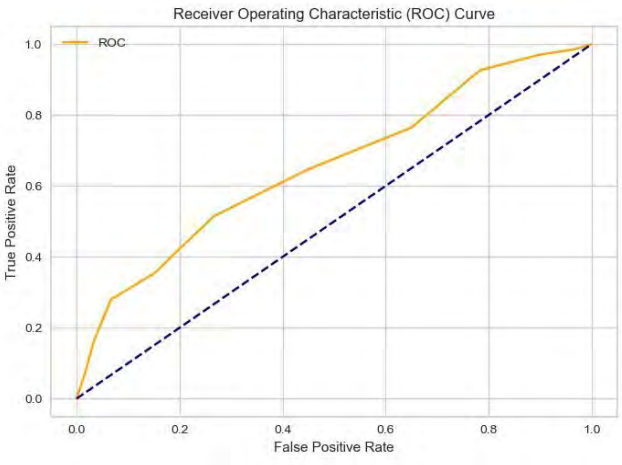
2. Each tree draws a random sample from the original data set when generating its splits, adding a further component of randomness that prevents overfitting.

In Machine Learning, performance measurement is an fundamental task. Therefore, when it comes to a classification problem, we can rely on an AUC - ROC Curve. At the point when we have to check or visualize the performance of the multi - class classification problem, we use AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve. It is one of the most significant evaluation metrics for checking any classification model's performance.

In the figure below, one can see the true positive rate vs the false positive rate for the random forest. We tested the algorithm for a particular dataset and it gave an accuracy of
78.67% which was the highest among the algorithms. Moreover, random forests can handle missing values very well. Although it takes slightly more time than the other algorithms, we decided to go for it because our primary focus is to make the prediction with maximum accuracy.

Apart from this, random forest also give the following advantages:-

- The predictive performance can compete with the best supervised learning algorithms
- They provide a reliable feature importance estimate
- They offer efficient estimates of the test error without incurring the cost of repeated



model training associated with cross-validation

III. PROPOSED WORK

When we started working on the project, we wanted to carry out live web scrapping to take the player data from the internet and use it for analysis so that we could make a prediction. At that time, our primary focus was speed as we wanted to make the prediction after every over of the second innings. Such a system would help us consider the records of matches directly. However, storing the dataset offline would mean that we would have to keep updating that data manually. A scrapping code would automate this. Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser. We used beautiful soup library to make sure that we could scrap the data. Although we tried to make the scrapping thing work, it was not possible. The problem with scrapping is that the websites do not allow scrapping of data very easily. Moreover, a slight change in the HTML code of the website could hamper the results of the data in a very significant way. Websites keep changing this code on a regular basis to prevent scrapping for taking place. Therefore, it becomes really difficult to carry out scrapping. A scrapping code which works right now may or may not work after two months. This could be a huge problem as we would need to change the code every now and then and that the code could stop giving the correct results at any point of time. Primarily, we were scrapping player data and team data from websites like ESPN Cricinfo and Cricmetric. Also, we were finding out the record of one batsman against all the bowlers of the opposition lineup. This data was going to be used to carry out the prediction of the winner of the match.

As you can see above, the user is asked to enter the teams that are playing in the match. Moreover, the date is to be entered as well. This is the date till which all the matches will be considered for making the prediction. The ground ID is the one on cricinfo website. This was making use of scrapping. However, we are now going to use the same model for making the prediction of IPL data. We have data of about 9 seasons which will be used to make the prediction. The UI will change accordingly. We will make a system which will predict the winner of the toss and their decision as well. Moreover, we will be able to predict the score that will be made by the team chasing and hence the winner of the game. After pre- processing and feature generating, we will use the machine learning to generate models which can be used to predict the results.

IV. CONCLUSION

With growing interest in the sport of cricket over the past few years, a need for tool which can predict the results of the match in advance has to be developed. Though there are lots of tools available in the market,

there accuracy lacks in the way they take factors into consideration. To give an edge to those existing classifiers, our project aims to also take into consideration some important factors like team composition, performance of players in past few matches, batting and bowling averages of the players in each team and winning probability of team batting first at a specific venue against a specific opponent. All these important factors along with toss and venue has taken into account and a classifier will be generated to give better results. Along with these a user interface has also been generated so that even a layman can interact with our system with ease.

The project currently takes into account the important factors from previous matches and then predicts the result before the match starts. But we can also take the ongoing match details to predict the result. So, the future work will be to combine both the previous data with ongoing match data to have much more better results. Also, the prediction can be extended to not only predict the winner of the match but also the expected runs to be scored by both teams.

Selection of the best team for a cricket match plays a significant role for the team's victory. The main goal of this paper is to analyze the IPL cricket data and predict the players' performance. Here, three classification algorithms are used and compared to find the best accurate algorithm. Random Forest is observed to be the best accurate classifier to predict the best player performance. This knowledge will be used in future to predict the best team.

References

[1] Ananda Bandulasiri Manage (2015). Predicting the Winner in One Day International Cricket, Sam Houston State University

[2] Ujwal U J et. At, "Predictive Analysis of Sports Data using Google Prediction API" International Journal of Applied Engineering Research", ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2814-2816.

[3] Swetha, Saravanan.KN (2017). Analysis on Attributes Deciding Cricket Winning.

[4] Passi, Kalpdrum & Pandey, Niravkumar. (2018) "Predicting Players' Performance in One Day International Cricket Matches Using Machine Learning" 111-126. 10.5121/csit.2018.80310.

[5] R. P. Schumaker, O. K. Solieman and H. Chen, "Predictive Modeling for Sports and Gaming" in Sports Data Mining, vol. 26, Boston, Massachusetts: Springer, 2010.

[6] Bunker, Rory & Thabtah, Fadi. (2017) "A Machine Learning Framework for Sport Result Prediction. Applied Computing and Informatics", 15. 10.1016/j.aci.2017.09.005.

[7] Monalisha Pattnaik, Anima Bag (2015).Fitting of Logistic Regression Model for Prediction of Likelihood of India Winning or Losing in Cricket Match.

[8] I. P. Wickramasinghe et. al, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Exercise", vol. 9, no. 4, pp.744-751, May 2014.

[9] Parag Shah, Mitesh Shah (2015).Predicting ODI Cricket Result.

[10] A Kaluarachchi, SV Aparna (2010).A classification based tool to predict the outcome in cricket.

Enhanced System for heart disease Prediction (Extended)

Jayshree Purohit, ShikhaYadav, Shivani Yadav, Prachi Janrao
Computer Engineering, TCET University of Mumbai
jayshree0198@gmail.com, shikhayadav2910@gmail.com, shivani19970114@gmail.com, prachi.janrao@thakureducation.org

Abstract

HeartDiseasePredictionSystem(Extended) (HDPS)isan websitethatpredicts occurrence ofHeartDiseasefrom medical datalikepainlocation, smoking, diabetes,whether pain relieved after rest, years, familyhistory using data miningtechnique.

Datamining techniqueplaysanimportant roleinthefield ofheartdiseaseprediction.Hence,therehasemergesaneed todevelopacomputerized systemfor checking the conditionofaheart-problem inapatient.Thuswehave includedaddition featuresinthesystemwhichcangivea newsortofperformancetothesystem,suchasdetailsabout thedoctorswhicharespecialist indealingwiththeissuesregardingtheheartproblemandt hesecondfeatureis t h a t ,too know about the hospitals were we could find the Doctors alongwiththeirfeesandbookingappointment whichisalsoavailableinthesystem.

Themainaimofthisprojectistoreducetheindividual and governmenthealthcareexpenditure.Thetraditionalmedi cal TestslikeMRIscan,ECG,CTscanarenotonlyexpensive Butalsorequiresheavymachinery andskilledoperatorsto operatethemachines. TheHeartDisease Prediction System (HDPS)willbeabletoreachalarge amount ofpopulation residing intheremoteareasthrough internetandwillalso eliminatethe heavymachineryandoperators.

Theprediction oftheHDPSwillbebasedonmachine learning algorithms. Thealgorithmwhichhashighest accuracywhile trainingon adatasetwill be chosen fordeploymentofthesystem.

Keywords: Cardiovascular Disease, DecisionTree, RandomForestClassifier,Logisticregression,KNN.

I. INTRODUCTION

HeartDiseasePrediction Systemwillhelpthepatientsto know about thestateofdiseasewhetherithasoccurornot. Notonly thepatientsbutalsothedoctors cangetaproper idea forhelpingthepatientswiththeirstateofdisease because the system will give an result in a form of Percentage, dependinguponthepercentage thedoctorscan work on.Asthisistheextended projectthepatients after getting theresults can even search forthehospital andthe doctorsdealingwiththissortofproblems.

II. METHODS AND MATERIAL

2.1 Existingsystem

Everyone isapatientatsomepointoftimeoranother,and weallwantgoodmedical care.Weassume thatdoctors Volume areallmedical experts andthatthereisgood research behindalltheirdecisions.However,thatcannot alwaysbethecase.Nevertheless, theycannotpossibly committomemory alltheknowledge theyneedforevery situation,andtheyprobablydonothaveitreadilyavailabl e.

Eveniftheexsistingsystemsdidhaveaccesstothemassiv e amountsofdataneededtocomparetreatment outcomesfor allthediseasestheyencounter, theywouldstillneedtime andexpertisetoanalysethatinformationandintegrateit with thepatient'sown medicalprofile. But thiskind of research andstatistical analysis isbeyondthescopeofa physician's work. Theywantadoctorwhowilltalktothem, listentowhattheysayandgivethemadviceabouthowto getbetterandprotecttheirhealth inthefuture.Inmany

cases, thewishforaprescriptionissecondarytothewish of beingcaredforDisadvantageofanexistingsystemwoul d bethatthepatientshavetovisitthedoctorinpersonandstil l doesnotgetpropertreatment, asthedoctorsareunableto predicttheexact disease. Humanerrorcan beavoidedwith thehelpofcomputer assistedquality decisionmaking. Itis timeconsuming whentherearehugeamounts ofdatatobe classified.Inaddition,efficiency andaccuracy ofdecisions willdecreasewhenhumans areputintostressandimmense work.Imagine adoctorwhohastoexamine fivepatientsrecords;he orshewillgo throughthem withease.However, ifthenumberofrecordsincreaseswithatimeconstraint,it isalmostcertain thattheaccuracy withwhichthedoctor delivers theresultswillnotbeashighastheonesobtained whenhe/shehadonlyfiverecordstobeanalysed.

II. PROPOSED SYSTEM

Toovercomethedrawbacksofthe existingsystemwe havedevelopedahealth predictionSystem.Thestructureof theproposedsystemisshowninFigurebelow:-

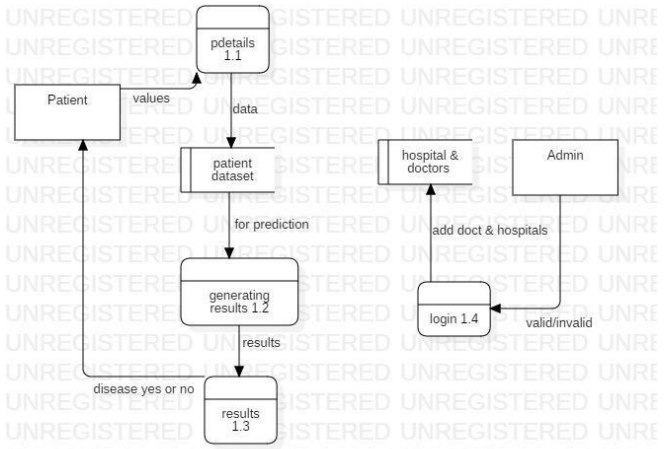


Fig3.1:DFDlevel1

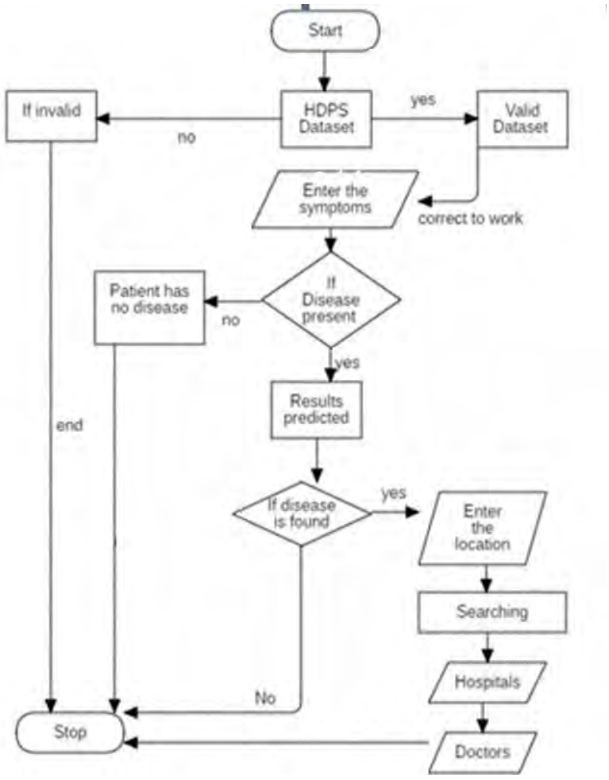


Fig3.2Flowchart

4 FeaturesofSystem

4.1 PatientModule

PatientRegistration:-ThePatients aresupposetoenterthe symthoms inthesystemforfindingtheresultdepending upontheresultsthefutherstepsareperformed.

4.2DoctorModule

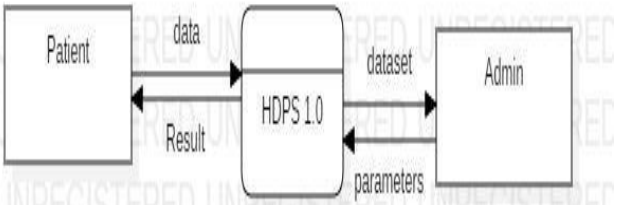
Doctor:-Thedoctorscan viewthereportofthepatients details whichcanhelpthemtotakethenecessary actions required.

4.3HospitalModule

TheHospital modulehelpsusto findthe nearby doctors alongwiththereareaof specialist

anditalsogivesa chancetomakeanappointment withthedoctorsfor discussingtheissues.

III. WORKING OF THE SYSTEM



According tothediagramsaformthatshowsalistof symptoms.Onthebasisofselectedsymptoms thesystem will generate related diseaseandbasisonthathewill searchnearbyhospitalanddoctor.

6.DataMiningTechniques

6. 1. KNearestNeighborsAlgorithm

- KNNisanon-parameterlearningalgorithm.The purposefor usingthisalgorithmis databasein whichthe datapointsare separatedinto several Classesforpredicting theclassificationofanew samplepoint.
- KNNworks on finding the distances between a queryandallthe**examples**inthedata,aswellas selectingthe specifiednumber of **examples**(K) closest to the query, then votes for the most frequentlabeloraveragesthelabels.

6. 2. RandomForest:-

- Randomforests algorithmis an ensemblelearning method for classification, regression and other tasksthatoperatesbyconstructing amultipleof decision treesattraining timeandoutputtheclass thatisthesourceoftheclasses ormean prediction oftheindividualtree.
- Randomforest isa termforan classifierthat consists ofmanydecisiontrees and outputs the classthatisthemodeoftheclasses outputby individualtrees.Randomforestsarecollectionsof trees,whichareallslightly different.Itrandomize thealgorithm,notthetrainingdata.

6.2.2 W o r k i n g ofRandomForestAlgorithm

The steps which will help us to execute the algorithmisas follows:-

- Step1− Theselectionofrandom samplesfroma givendataset.
- Step2− t h e algorithmwillhelpin c o n s t r u c t i n g adecisiontreeforeverysample.Thentheresult willbepredicteddependinguponoutputofrstof thedecisiontrees.
- Step 3− This step contains , voting will be performedforeverypredictedresult.
- Step4–The lastinthealgorithmis,selectthe mostvotedprediction resultasthefinalprediction result.

Acknowledgment

We sincerelythankour guideMs.PrachiJanraoforher guidance andconstantsupport andalsoforthesticktoour backs.Wealsothankstheprojectcoordinators forarranging the necessaryfacilitiestocarryouttheprojectwork.

WethanktheHOD,Dr.SheetalRathi,DeanAcademi c, Dr. R.R Sedamkar,the Principal,Dr. B. K. Mishra andthe collegemanagementfortheir support.

References

- 1.DiseasePredictionUsingHeartrateVariability Analysis-IoTDharmikJampala1,VenkatNaidu Mittapalle2, SitanshuNandan31,2,3Schoolof InformationTechnology andEngineering,VIT University,Vellore.
- 2.K.Vembandasamy,R.Sasipriya,andE.Deepa, "HeartDiseases Detection UsingNaiveBayes Algorithm",IJISET-International Journalof InnovativeScience,Engineering &Technology, Vol.2,pp.441-444,2015.

3.Sanjay KumarSen, IJECS Volume 6 Issue 6 June,2017PageNo.21623-21631

- 4.Animesh Hazra, ArkomitaMukherjee, Amit Gupta,Asmita Mukherjee, "HeartDisease Diagnosis andPrediction UsingMachineLearning and Data MiningTechniques:A Review", ResearchGatePublications,July2017,pp.2137- 2159.

- 5.V.V. Ramalingam, AyantanDandapath, M KarthikRaja,"Heartdiseaseprediction using machinelearningtechniques: asurvey", InternationalJournal of Engineering &Technology,7 (2.8)(2018)684-687.
- 6.AnimeshHazra,SubrataKumarMandal,Amit Gupta, ArkomitaMukherjee and Asmita Mukherjee, "HeartDiseaseDiagnosis and Prediction UsingMachineLearningandData MiningTechniques: AReview",Advances in Computational Sciencesand Technology ISSN 0973-6107Volume10,Number7(2017)pp. 2137- 2159.
- 7. Asha Rajkumarand B. Sophia Reena, "DiagnosisofHeartDiseaseUsingData miningAlgorithm", GlobalJournalofComputerScience and Technology,Vol. 10, No. 10, pp. 38 -43,2010.
- 8.Animesh Hazra, ArkomitaMukherjee, Amit Gupta, AsmitaMukherjee, "Heart Disease Diagnosis and Prediction UsingMachine Learning and Data Mining Techniques: A Review", ResearchGate Publications, July2017,pp.2137-2159.
- 9.V.Krishnaiah, G. Narsimha, N. Subhash Chandra, "Heart DiseasePrediction System using Data Mining Techniques andIntelligent Fuzzy Approach: A Review", InternationalJournalofComputerApplications, February2016.
- 10.GuizhouHu, Martin M. Root, "Building Prediction Models for Coronary Heart Disease by Synthesizing Multiple Longitudinal Research Findings", European Scienceof Cardiology,10May2005.

Object Sorting based on Color Sensing using Mat Lab

Deepak Jaiswar Ramzan khan, Rishabh Dubey, Mrs..Kalpana Gangwar
Dept. Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
deepakjaiswar120@gmail.com ramzank268@gmail.com dubey.rishabh009@gmail.com Kalpana.iiitm@gmail.com

Abstract— ***Article arranging is utilized in many bundling industry at the hour of definite dispatch .In these day manual item arranging is a conventional methodology favored by ventures. Consequently decided endeavor have been done to create and structure programmed procedure which can naturally sort and check object dependent on shading sensor and picture preparing instrument. Through shading, we can approve, track separate various articles in progressively better way. Color sensor is utilized to recognize shade of item. This strategy can be used to recognize moving items on transport line and can be recognize by shading sensor. Shading is the most well-known component to recognize between objects, arranging, perceiving and following. For the most part robot is mounted with a camera or the camera is mounted in the workspace to distinguish the article. This innovation can be utilized in material dealing with in coordination's and bundling industry where the articles traveling through a transport line can be isolated utilizing a shading identifying robot. Right now 'Object' calculation is written in MATLAB for playing out the activity. The 'Object' calculation is executed to distinguish the question and send the proper directions to the microcontroller utilizing sequential correspondence for the robot to play out the arranging activity***

Keywords: Mat lab, Color Object Counting, Image Processing, Object Sorting and Adriano Uno

INTRODUCTION

Arranging of items is a troublesome mechanical Procedure. Nonstop manual arranging makes consistency Issues. This undertaking depicts a working model intended for programmed arranging of articles dependent on the Shading. TCS230 sensor is utilized to identify the shade of the item and the Adriano Uno MATLAB are utilized to control the general procedure. The recognizable proof of the shading depends on the recurrence examination of the yield of TCS230 sensor. Two transport lines are utilized, each constrained by independent DC engines.

The principal belt is for setting the item to be broke down by the shading sensor, and the subsequent belt is for moving the holder, having isolated compartments, so as to isolate the items. The trial results guarantee that the model will satisfy the requirements for higher creation and exact quality in the field of mechanization. Machines can perform profoundly dull undertakings superior to people. Laborer exhaustion on sequential construction systems can bring about decreased execution, and cause difficulties in keeping up item quality. A representative who has been playing out an examination task again and again may in the long run neglect to perceive the shade of item. Mechanizing a significant number of the undertakings in the enterprises may assist with improving the proficiency of assembling framework. The motivation

III. EXISTING SYSTEM

In the current framework the items are arranged physically for the most part by individuals. This makes a propensity for human mistakes to come into record and consequently bring about the work turning out badly. On the off chance that articles or parts in ventures are not arranged accurately, at that point there is item being blemished. a high possibility of tremendous turmoil and the last

IV.PROPOSED SYSTEMS

The disadvantages of existing systems can be overcome using the proposal systems. In the proposed system a color sensor will be mounted on top of the conveyor belt along with a proper support. The color sensor will detect the color of the object based on the frequency and will send this information to Mat lab. This will help to keep track of the objects of a particular color and also record when they were detected.

V. TECHNICAL OVERVIEW

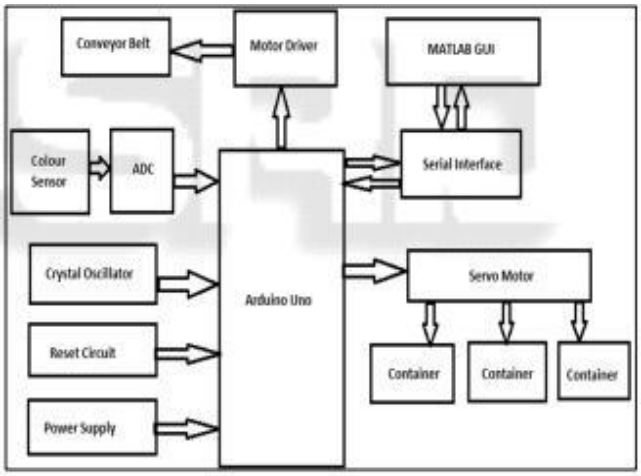


Fig. 1: Diagrammatic representation

The main objective of the system being developed is to detect the object which is coming from the conveyor belt with the help of a TCS230 color sensor and sort the object based on their color. The conveyor belt is being driven by a motor which is connected to the Adriano Uno via a driver as the voltage required by the motor is more than what the Adriano Uno can provide. The frequency of color which is detected by the color sensor is sent to the Adriano Uno which in turn tells if the object has to be separated or not based on the programming done in the Adriano Uno software installed on the computer. The GUI is developed with the help of Mat lab in order to display the object detected and sorted. Thus Mat lab plays a key role over here. Some LED's are mounted upon the Adriano Uno which will blink according to their respective colors when the object of that color is detected. Object Sorting System based on Color Sensing using Mat Lab (IJSRD/Vol. 4/Issue 11/2017/068) All rights reserved by www.ijsrd.com 243

VI. DESIGN OF THE SYSTEM

Design of the system will include: 1) Adriano Uno: The Adriano Uno is a microcontroller board based on the ATmega328 (datasheet). It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button. 2) Color Sensor: The Color Sensor is a complete color detector. It consists of a TAOS TCS3200 RGB sensor chip and 4 white LEDs. It can detect and measure a nearly limitless range of visible colors to a certain degree. 3) Dc Motor: A DC motor is any of a class of rotary electrical machines that converts direct current electrical power into mechanical power. The most common types rely on the forces produced by magnetic fields. Nearly all types of DC motors have some internal mechanism, either electromechanical or electronic, to periodically change the direction of current flow in part of the motor. 4) Motor Driver: A motor driver is a little current amplifier; the function of motor drivers is to take a low-current control signal and then turn it into a higher-current signal that can drive a motor. 5) Mat lab: MATLAB (matrix laboratory) is a multiparadigm numerical computing environment and fourth-generation programming language. A proprietary programming language developed by Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran and Python.

VII. TECHNOLOGY USED

Why Mat lab Instead of VB? Code written right now effectively comprehensible. Mat lab has more highlights when contrasted with Visual Basic. This language exceeds expectations at simultaneousness when contrasted with Visual Basic. Mat lab has a decent Library Distribution instrument when contrasted with Visual Basic. Mat lab is entirely appropriate for Real Time Applications. Mat lab has a wide scope of implicit Pre-Written codes which make the undertaking of coding less troublesome and efficient. Fig. 2: Mat lab Software

VIII. ADVANTAGES OF MATLAB

Mat lab is an easy to use and effectively reasonable language. Codes written right now to be entirely solid. Mat lab is helpful for numeric and logical processing. Its essential information component is the framework. A straightforward whole number is viewed as a grid of one line and one section. Vector zed tasks. The graphical yield is upgraded for collaboration. Mat lab's usefulness can be enormously extended by the expansion of tool stash. An enormous (and developing) database of implicit calculations for picture preparing and PC vision applications MATLAB permits you to test calculations promptly without recompilation. You can type something at the direction line or execute a segment in the supervisor and quickly observe the outcomes, extraordinarily encouraging calculation improvement. The MATLAB Desktop condition, which permits you to work intuitively with your information, encourages you to monitor records and the factors and improves regular programming troubleshooting undertakings the capacity to peruse in a wide assortment of both normal and area explicit picture designs. The capacity to call outer libraries, for example, OpenCV.

IX. HISTORY OF MATLAB

Cleve Moller, the chairman of the computer science department at the University of New Mexico, started developing MATLAB in the late 1970s. He designed it to give his student's access to LINPACK and EISPACK without them having to learn FORTRAN. It soon spread to other universities and found a strong audience within the applied mathematics community. Jack little, an engineer, was exposed to it during a visit Moller made to Stanford University in 1983. Recognizing its commercial potential, he joined with Moller and Steve Bangers. They rewrote MATLAB in C and founded Math Works in 1984 to continue its development. These rewritten libraries were known as JACKPAC. In 2000, MATLAB was rewritten to use a newer set of libraries for matrix manipulation, LAPACK. MATLAB was first adopted by researchers and practitioners in control engineering, Little's specialty, but quickly spread to many other domains. It is now also used in education, in particular the teaching of linear algebra, numerical analysis, and is popular amongst scientists involved in image processing.

X. MATLAB FEATURES

Significant level language for logical and designing registering. Work area condition tuned for iterative investigation, structure, and critical thinking. Designs for imagining information and devices for making custom plots. Article Sorting System dependent on Color Sensing utilizing Mat Lab (IJSRD/Vol. 4/Issue 11/2017/068) All rights held by www.ijsrd.com 244 Apps for bend fitting, information grouping, signal examination, and numerous other space explicit errands. Extra tool compartments for a wide scope of building and logical applications. Instruments for building applications with custom UIs. Interfaces to C/C++, Java®, .NET, Python®, SQL, Hardtop®, and Microsoft® Excel® without royalty arrangement choices for sharing MATLAB programs with end clients.

XI. REQUIREMENT ANALYSIS

A.Possibility Study

The possibility study assumes a significant job in the investigation of the framework. The very choice of the framework expert, to structure a specific framework depends whether the framework is attainable or not. Henceforth, the attainability study shapes the essential of the framework. The plausibility study can be ordered into: 1) Technical Feasibility It has been resolved that the innovation required for the arranged framework is open which this innovation might be coordinated into the apparatus. Specialized examination has conjointly assessed the present framework to search out that it can't be redesigned keep with the client's needs. Consequently, we'd prefer to make a total new framework that obliges the specific needs of the client. The tip client might be outfitted with the pre-referenced equipment and code necessities. 2) Operational Feasibility There are two parts of operational plausibility for the framework. One angle is that of specialized execution and other is of Acknowledgment. It has been resolved that the framework can give right and convenient information required for the end client needs. Additionally it has been resolved that the framework will be acknowledged by the two clients with and without specialized information. 3) Economic Feasibility The financial practicableness of the framework is primarily engaged with its cash angles. It decides if the undertaking is financially conceivable. As the equipment and PC code are as of now out there basically inside the market, no greater speculation is to be made toward that path, the sole cost concerned is that of executing the framework. The side framework utilizes mat lab that is uninhibitedly out there. The backend is worked in c coding that the is put forth out there with clear attempts and lower cost.

XII. MATLAB INTEGRATES WORKFLOWS

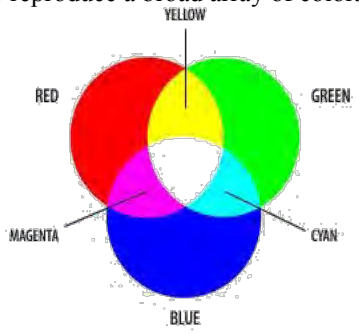
Major engineering and scientific challenges require broad coordination to take ideas to implementation. Every handoff along the way adds errors and delays. MATLAB automates the entire path from research through Production. You can: Build and package custom MATLAB apps and toolboxes to share with other MATLAB users. Create standalone executable to share with others who do not have MATLAB. Integrate with C/C++, Java, .NET, and Python. Call those languages directly from MATLAB, or package MATLAB algorithms and applications for deployment within web, enterprise, and production systems. Convert MATLAB algorithms to C, HDL, and PLC code to run on embedded devices. Deploy MATLAB code to run on production Hardtop systems. MATLAB is also a key part of Model-Based Design, which is used for multi domain simulation, physical and discrete-event simulation, and verification and code generation. Explore Simulink, Sims cape™, and State flow® to learn more about Model-Based Design.

XII. HARDWARE AND SOFTWARE REQUIREMENT

A. Hardware Requirement
Adriano Uno
Conveyer belt
Power supply
TCS230 color sensor
Dc motor
Servo motor
B. Software Requirement
-Mat lab

XIII. RGB COLOUR MODEL

The RGB color model is a color model in which consists of red, green and blue light are added together in various ways to reproduce a broad array of colors.



2. HSV Color MODEL The HSV shading space has the accompanying three parts: H – Hue (Dominant Wavelength). S – Saturation (Purity/shades of the shading). V – Value (Intensity). Shade (H) characterizes the shading itself. Immersion (S) is the measure of dim from zero percent to 100 percent in the shading. The worth (V) Component catches the measure of light falling on it in this way it changes because of brightening changes. HSV utilizes just one channel to depict shading, making it extremely instinctive to determine shading as appeared in figure.

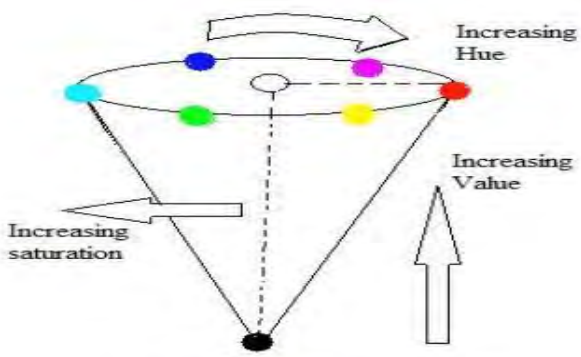


Figure-2 HSV Colour model

XIV. SERVOMOTOR

Servomotors are used in applications such as robotics, CNC machinery or automated manufacturing. The servo motor has some control circuits and a potentiometer (a variable resistor)

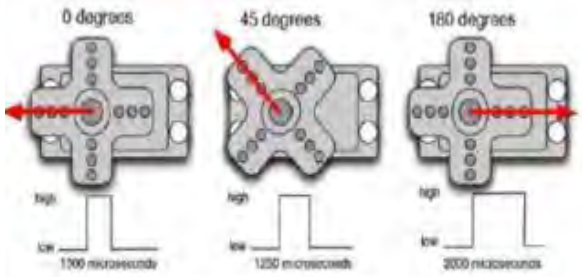


Figure 4. Servomotor Rotation

XIV. CONCLUSION

The conclusions drawn from results given by algorithms used for automatic cooler object counting and sorting in prototype system design to implement automation in automatic technique to determine cooler of object ,object count and sort object based on color using image processing approach are as follows, An approach for object cooler detection, count calculation and object sorting has been implemented. Implemented system gives accurate result for purely Red, Green and blue colored objects. With some software changes this system can be used for different shades of basic specified cooler. Due to use of automation in cooler determination and count calculation process, manual Pot allows the control circuitry to monitor the current angle of the servo motor. If the shaft is at the correct angle, then the motor shuts off. If the circuit finds that the angle is not correct, it will turn the motor the correct direction until the angle is correct. The output shaft of the servo is capable of traveling somewhere around 180 degrees. Usually, it's somewhere in the 210 degree range, but it varies by manufacturer. A normal servo is used to control an angular motion of between 0 and 180 degrees. If the output position differs from that required, an error signal is generated which then causes the motor to rotate in either direction, as needed to bring the output shaft to the appropriate position. As the positions approach, the error signal reduces to zero and the motor stops. Servomotors measure both the position and also the speed of the output shaft. They may also control the speed of their motor, rather than always running at full speed. Both of these enhancements, usually in combination with a PID control algorithm, allow the servomotor to be brought to its commanded position more quickly and more precisely, with less overshooting. The servo turn rate, or transit time, is used for determining servo rotational velocity. This is the amount of time it takes for the servo to move a set amount, usually 60 degrees. The servomotor rotation is shown in figure 4.For example, suppose you have a servo with a transit time of 0.17sec/60 degrees at no load, this means it would take nearly half a second to rotate an entire 180 degrees.

efforts are reduced which leads to improving accuracy as well as saves money and time.

Xv. References

[1] Vindhya Devalla, Dr. Rajesh Singh, Amit Kumar Mondal, Vivek Kaundal, "Design And Development of

Object Recognition And Sorting Robot For Material Handling In Packaging and Logistic Industries", International Journal of Science and Advanced Technology (ISSN 2221-8386) Volume 2 No 9 , pp 30-35,September 2012

[2] Vishnu r. Kale, V. A. Kulkarni, "AUTOMATION OF OBJECT SORTING SYSTEM USING PICK & PLACE ROBOTIC ARM & IMAGE PROCESSING" ,

Proceedings of 3rd IRAJ International Conference, 5th January 2014, Mumbai, India. ISBN: 978-93-82702-51- 1 Object Sorting System based on Colour Sensing using MatLab (IJSRD/Vol. 4/Issue 11/2017/068) All rights reserved by www.ijsrd.com 245

[3] Aashik Chandramohan, Krishna Murthy K.R., Sowmya G., Surya Prasad P.A., Vijay Krishna V., Peeyush K.P., "Financially savvy Object Recognition and Sorting Robot utilizing Embedded Image Processing Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-3, Issue-11,pp-29-32, April 2014.

[4] C. Chandra Mouli, P. Jyothi, K. Nagabhushan Raju "Displaying AND IMPLEMENTATION OF WIRELESS EMBEDDED ROBOT ARM FOR OBJECT SORTING", IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) e-ISSN: 2278-1676,p-ISSN: 2320-3331, Volume 9, Issue 4 Ver. III (Jul – Aug. 2014), PP 35-44

[5] Atef. A. Ata, Sohair F. Rezeka, Ahmed El-Shenawy, Mohammed Diab "DESIGN AND DEVELOPMENT OF5-DOF COLOR SORTING MANIPULATOR FOR INDUSTRIAL APPLICATIONS", International Journal of Mechanical, Aerospace, Industrial and Mechatronics Engineering Vol:7, No:12, 2013

Movie Release Scheduling Using Machine Learning

Sakshi Singh, Sruthi Harish, Yashesh Oza
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India s
akshidsingh000@gmail.com, sruthiharish11@gmail.com, yashesh654@gmail.com

Abstract— **The scheduling of movie screenings in the multiplexes these days is heavily based on the managers' gut feeling. Movie theatres can maximize their profits using a model that analyzes a lot of factors like the genre of the film, it's budget, the popularity of the lead actors, the estimated revenue, etc. to predict a near to optimal outcome. The main aim proposed is to screen a movie based on the availability of the targeted audience and to maximize the predicted revenue. The prediction of the revenue of the movie also depends on a variety of factors like its budget, popularity, associated production companies and most importantly release date. To release a movie in the perfect time is to maximize the revenue estimate. This makes it necessary to reform the existing system of manual movie scheduling and replacing it with a movie forecasting model that learns from previous data. However, the limitations are as follows: Firstly, considering the complexity of movie forecasting, it becomes excessively difficult to carry out micro-scheduling which deals with the internal distribution of movie screenings both with respect to space and time of these and hence, we may settle for macro-scheduling which describes a model which specifies the optimal dates for screening the movie. Secondly, a few more factors may add to the perfect time for the release of the movie, like the weather and the location of a multiplex. These cannot be fixed and would vary. To understand the significance of the movie release dates on the revenue generated, we have to gauge their dependence using the optimized and powerful decision tree-based ensemble Machine Learning model of XGBoost that uses a gradient boosting framework.**

Keywords—Movie Forecasting, Micro-scheduling, Macro- scheduling, Decision Tree , Ensemble Based, XGBoost, Gradient Boosting Framework

I. INTRODUCTION

1.1 Introduction

Living in a completely tech-era, where almost every stuff we see daily is mostly automated, the scheduling of movie releases remains to be one of those tasks which require manual effort. The manual effort required proves to be frustrating, cumbersome and most importantly does not guarantee to yield optimal results. Optimality is a large scope here in terms of the targeted audience, revenue generated, etc. All of these though does not singlehandedly depend on the release date of the movie, but the release date is one of those variable factors which on an appropriate modification, can push up the revenue generated. Hence by varying the release date based on the availability of the targeted audience, would surely help in strategizing an appropriate date for the release of the movie. As proved ahead, the only feasibly varying component on which the revenue generated depends largely on is the release date of the movie. Taking an example to support this claim, an animation movie meant to entertain the kids would do no good to the revenue if it is screened at a time when the kids are not free(for example during exams, or during weekdays when they have schools). Going in-depth, it is understood that each movie has a targeted audience, even though it tries to capture all age- groups equally. The targeted audience depends on factors like the genre of the movie, its lead actors, its directors, production house, etc. These are pretty repetitive which would help to train the model based on this data helps to learn out of the experience

1.2 Motivation

Forecast model used to assist film studios as even a single movie can be the difference between crore of rupees of profit or loss in a given year. Forecast model also used to assist cinema hall/multiplex owners in planning out movie schedules for forthcoming box-office weekends. Forecast model will let the people effectively schedule the movies in less time without doing any manual calculations. This model will find out the target audience for a particular movie so that the profit will be more.The current existing system for scheduling movies is done manually with pen & paper by specialists. Despite the combination of a programmer's analytical mind & a broad knowledge about movies & audience of individual theatres, an analytical model can help do better

1.3 Literature Survey

Is about a marketing decision support system in the movie industry. The decision support system of interest is a model that generates weekly movie schedules in a multiplex movie theater. A movie schedule specifies, for each day of the week, on which screen(s) different movies will be played, and at which time(s). The model integrates elements from marketing with approaches from operations research .Therefore, it consists of two parts: (i) conditional forecasts of the number of visitors per show for any possible starting time, and (ii) a scheduling procedure that quickly finds a near optimal schedule .[2] proposes the use of a review system for predicting the success rate of a movie. Moviegoers' opinions of a movie before and after the release of the movie will be determined using sentiment analysis. [3] introduces the concepts of

neural networks and presentsan overview of the applications of neural networks in decision support systems (DSS).[4] suggests that movie theatres could improve profits by using a more data- driven approach such as movie attendance forecasting model. Eilashberg used linear regression model, standard logit model and nested logit model. [5] proposes a DSS for movie investment sector using machine learning techniques. It proposes the use of Support Vector Machine(SVM), Neural Network(NN) and Natural Language Processing(NLP).

1.4 Problem Definition

There is a heavy interdependence between the movie release date and the revenue it generates. This is possibly the only feasible and variable factor(that can be easily changed without creating much complications) which influences the revenue. The major task is to maximize the revenue generated by optimally fixing the movie's release date. This though sounds simple, depends a lot of other factors.

1.5 Problem Solution

The aim is to forecast the Box-office revenue for a film using Machine Learning algorithm Decision Tree,by adding the computed influence of each parameter of a movie that is believed to affect its revenue.In this project we also aim to find the target audience for the movie and also aiming to effectively schedule the movie so that the revenue generated will be maximum.

1.6 Proposed System

We propose a system that would help to efficiently schedule the movies to be released in the theatres such that the entertainment revenue generated is optimized. Different genres of movies are meant to entertain a different audience altogether. To increase the revenue generated, it would be best to release the movie when the targeted audience is free. For example, an animation movie is meant to entertain the kids of age group ranging from 3 to 8, or action-packed film is more likely to entertain boys of age group say 3 to 12, etc. So releasing these movies at the time of school exams would in no way help in optimizing the revenue. Apart from these, it is also necessary to consider certain other factors like the location of the multiplex, the actors' popularity, etc. These together in a combined way determine the popularity of the movie among the targeted crowd. The more the popularity, the more is the audience which automatically increases the revenue collected.

1.7 Objectives

System to efficiently schedule the movies to be released in the theatre

System to learn the target audience based on various features like the actors genres of the movie Schedule the movies as per the availability of the targeted audience Optimize the entertainment revenue generated Elimination of the current manual (pen & paper) approach.

1.8 Scope

The scheduling of movie screenings in the multiplexes these days is heavily based on the managers' gut feeling. Movie theatres can maximize their profits using a model that analyzes a lot of factors like the genre of the film, it's budget, the popularity of the lead actors, the estimated revenue, etc. to predict a near to optimal outcome. The main aim proposed is to screen a movie based on the availability of the targeted audience and to maximize the predicted revenue. The prediction of the revenue of the movie also depends on a variety of factors like its budget, popularity, associated production companies and most importantly release date. To release a movie in the perfect time is to maximize the revenue estimate. This makes it necessary to reform the existing system of manual movie scheduling and replacing it with a movie forecasting model that learns from previous data.

1.9 Application

Living in a completely tech-era, where almost every stuff we see daily is mostly automated, the scheduling of movie releases remains to be one of those tasks which require manual effort. The manual effort required proves to be frustrating, cumbersome and most importantly does not guarantee to yield optimal results. Optimality is a large scope here in terms of the targeted audience, revenue generated, etc. All of these though does not singlehandedly depend on the release date of the movie, but the release date is one of those variable factors which on an appropriate modification, can push up the revenue generated. Hence by varying the release date based on the availability of the targeted audience, would surely help in strategizing an appropriate date for the release of the movie. As proved ahead, the only feasibly varying component on which the revenue generated depends largely on is the release date of the movie. Taking an example to support this claim, an animation movie meant to entertain the kids would do no good to the revenue if it is screened at a time when the kids are not free(for example during exams, or during weekdays when they have schools).

1.10 Impact

Going in-depth, it is understood that each movie has a targeted audience, even though it tries to capture all age- groups equally. The targeted audience depends on factors like the genre of the movie, its lead actors, its directors, production house, etc. These are pretty repetitive which would help to train the model based on this data helps to learn out of the experience. So, obtaining the past data to determine the future most probable targeted audience would help in mapping a movie to the most optimal release dates. The project

would help in better scheduling of release dates which contributes to the revenue generated, helping in maximizing it.

1.11 Organization

Basically, the first step of the project involves understanding and assessing the interdependence between the scheduling and revenue generation. Once this, is completed the scheduling of movie releases as per the availability of the targeted audience is to be done. The past data of movies which has a lot of components like lead actors, production house, director, genres etc. is to be used to get appropriate and optimal results.

1.12 Expected Outcome

The expected outcome of our project is to find out the effective cost for the movie so that the revenue generation will be maximum and also expecting to find out the exact target audience for a particular movie based on the genre of the movie and to effectively schedule the movie in the film industry so that we can get more profit from movie based on weekly basis.

1.13 Limitations

Limitations in this domain are as follows: Firstly, considering the complexity of movie forecasting, it becomes excessively difficult to carry out micro-scheduling which deals with the internal distribution of movie screenings both with respect to space and time of these and hence, we may settle for macro-scheduling which describes a model which specifies the optimal dates for screening the movie. Secondly, a few more factors may add to the perfect time for the release of the movie, like the weather and the location of a multiplex. These cannot be fixed and would vary.

II. METHODOLOGY

A. Workflow

Firstly, to prove that there is a vast interdependence between the release date of the movie to the revenue, we design a model which predicts the revenue of the movie, doing so, using the optimized and powerful decision tree- based ensemble Machine Learning model of XGBoost that uses a gradient boosting framework. This model would help us gauge the important of various features in deciding or rather, maximizing the entertainment revenue.
Further, once it is proved that the release date of a movie stands out as one of the most important features to predict a movie's revenue generated, we would move on to the part of movie forecasting that is, its release date scheduling using the below demonstrated workflow

Figure 1 is a self-explanatory figure that demonstrates the general approach to solve the problem.

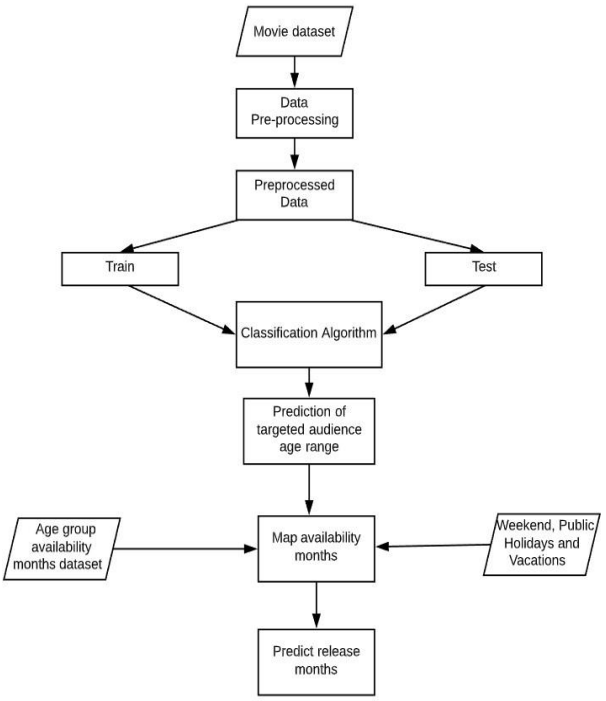


Fig 1: Workflow/ Work Methodology

III. RESULTS AND DISCUSSION

To prove that there is a vast interdependence between the release date of the movie to the revenue, we design a model which predicts the revenue of the movie, doing so, using the optimized and powerful decision tree-based ensemble Machine Learning model of XGBoost that uses a gradient boosting framework. This model would help us gauge the important of various features in deciding or rather, maximizing the entertainment revenue.
Dataset source: Kaggle Environment used: Google Colaboratory

Following are the snapshots of working as well as outputs obtained:

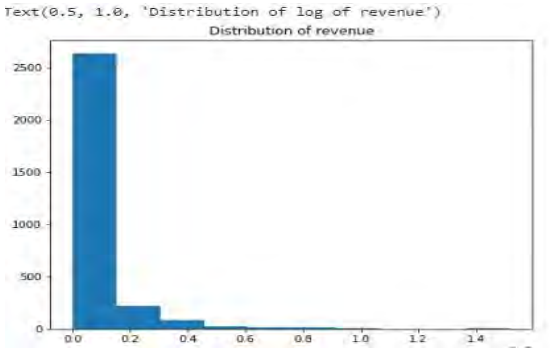
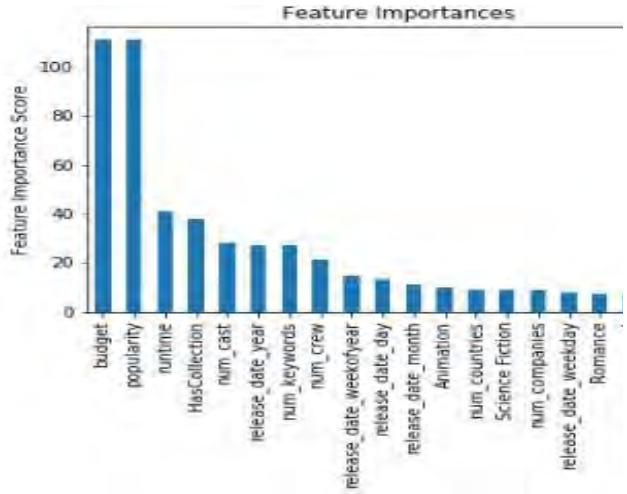


Fig 2: Distribution of Revenue

Figure 2 demonstrates the distribution of revenue .



The shaded region in the form of histogram graph exhibits the abundance of movies lying in that particular revenue scale.

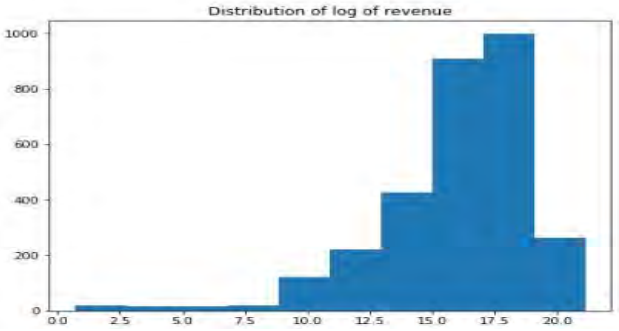


Fig 3: Distribution of log of revenue

Figure 3 demonstrates the distribution of logarithmic revenue.

The histogram graph includes the shaded region that specifies the abundance of movies in the dataset lying in the paritcular logarithmic of revenue scale.

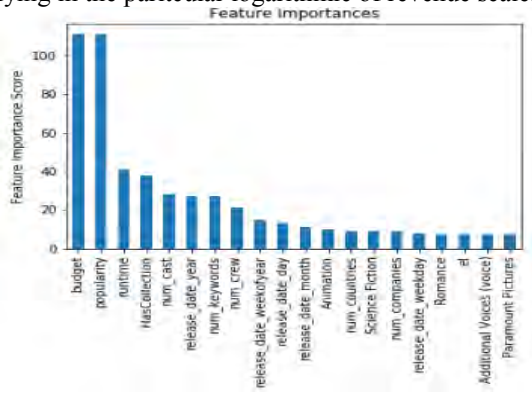


Fig 4: Feature Importance Score Graph

Figure 4 is one of the most important conclusions we draw from the dataset which governs the core of this research paper.

Feature Score graph points out how heavily or lightly would the entertainment revenue generated from the movie released depend on the corresponding factor or feature present along the x-axis of the given graph.

Relative % = Relative probability of dependence * 100%

The histogram graph has along y axis the relative percentage of dependence. As probability lies in range [0,1], relative percentage lies in the range of [0,100].

As per the Figure 4, budget and popularity of the movie heavily determine its revenue generated. That is, if a movie is more popular and its budget is towards the higher end , its more probable to yield a higher revenue.

The runtime of the movie as well as its cast affect the movie's revenue. But out of all of these, one of the easily variable factors that decide the revenue of a movie heavily is its release date. Hence, it is very evident that manipulation of the release date of the movie has a heavy impact on the revenue generated from the released movie.

IV. CONCLUSION

From the above implementation it is clearly seen that the revenue depends not only the release date but also the year, day and month. In fact it depends on the mere fact whetheher its released on weekend or a weekday. Such a small thing can make a vast difference. But the best part is, this can be planned and released optimally to give the best desired results.

V.ACKNOWLEDGEMENTS

- We sincerely thank our guide Dr. R. R. Sedamkar for his guidance and constant support and also for the stick to our backs.
- We also thank the project coordinators for arranging necessary facilities and resources and supporting and promoting among us the culture of research necessary to carry out the required research.
- We thank the Head of the Department, Dr. Sheetal Rathi, the Principal, Dr. B.K. Mishra and the college management for their support.

References

[1] Jehoshua Eliashberg, Quintus Hegie , Jason Ho, Dennis Huisman , "Scheduling of movie in a multiplex,",2009.

[2] Jim Hee Jinny Lim, A Thesis on Forecasting Movie Attendance of Individual Movie Showings: A Hierarchical Bayes Approach,Simon Fraser University,2012.

[3] Antara Upadhyay, Nivedita Kamath, Shalin Shanghavi, Tanisha Mandvikar, Pranali Wagh "Movie Success Prediction using Data Mining," in IJEDR, vol. VI,Issue IV, Shah & Anchor Kutchhi Engineering College, Mumbai: Academic, pp. 198–203.

[4] N. Quader, Md. O. Gani, D. Chaki, Md. H Ali, "A Machine Learning Approach to Predict Movie Box-Office Success", in 20th International Conference of Computer and Information Technology (ICCIT), 22-24 December, 2017.

[5] Javaria Ahmad,Prakash Duraisamy,Amr Yousef,Bill Buckles, "Movie Success Prediction Using Data Mining", in 8th International Conference On Computing, Communication And Networking Technologies (ICCCNT),3 -5 July, 2017.

Abstractive Text Summarization Using Deep Learning Approach

Saloni Shetty, Aakash Pandey, Harsh Rathore, Mrs. Vaishali Nirgude

Department of Computer Engineering

Thakur College of Engineering and Technology Mumbai, India

saloniashetty@gmail.com, aakashpande201@gmail.com, harsh.rathore1298@gmail. Com, vaishali.nirgude@thakureducation.org

Abstract - **Text summarization is an approach to summarize the massive amount of information into an abridged form by the process of selection of substantial information and discarding immaterial and redundant information. There are many applications where summary of large documents such as news headlines, market review, short messages on mobile, research work, medical need such as patients history for further treatment, business analysis etc. are needed. Text summarization can be classified into Extractive Summarization and Abstractive Summarization. An extractive summary involves selection of significant sentences from the original documents based on different attributes or features of sentences. Abstractive Summarizationprovidescrisp, efficient and more human-like summary but due to its complexity, much research hasn't been conducted in its domain. The paper aims to depict the usage of RNN and ConceptNetNumberbatch tool for Word Embedding to create powerful summaries of the product reviews obtained from any E-commerce websites .**

Keywords:NeuralNetworks,TextSummarization,Abstractive Summary, Deep Learning, Natural Language Processing

I. INTRODUCTION

A. NEED

In the current scenario where enormous information is available on the internet, it is most important to devise an intuitive approach in order to extract the apt information in a quick and efficient manner. As it is a tedious task for human race to manually extract the summary of large documents of text, so the need of Automatic Text Summarizer arises. Text summarization is the process of recognizing the premier significant information in a document or set of related documents and transforming it into a compact version such that the overall meaning of the original text isretained.

B. RELATEDTHEORY

Synopsis of a huge text is generated using text summarization techniques which is a sub domain of Natural Language Processing(NLP). Text can be summarized using two simple methods, they are NLP based method and deep learning based method.

The popularity of Automatic Text Summarization rose in the early 1950's. Hans Peter Luhn published a paper in 1958 titled "The automatic creation of literature abstracts". To obtain meaningful sentences from the literature he utilized methods like word and phase frequencies for summarization. In 1969 Harold P Edmundson carried out a few significant researches in the domain of automatic text summarization. To extract influential sentences for text summarization he used features such as the presence of words used in the title appearing in the text, location of the sentences, and cue words. Post this,the topic garnered a lot of attention in the field of Research & Development.

Later developments were greatly influenced by the rise of Machine Learning techniques in 1990's that used different techniques to create literature abstracts. Neural Network and common words in search engine queries are emerging methods to further improvise document summarization.Some of the technologies commonly used to create efficient summaries are:

i. *Deep learning*:

Neural Networks (NN) are also utilized for Natural Language Processing (NLP), including Summarizers. Neural networks are viable in practically solving almost any machine learning classification problem. Deep learning uses deep neural networks to learn good representations of the input data, which would then be utilized to perform specific tasks.

ii. Long Short Term Memory (LSTM)Units

The LSTM is a RNN architecture that can remember past contextual values. These stored values do not modify over time while training the model. There are four components in LSTM which are LSTM Units, LSTM Blocks, LSTM Gates and LSTM Recurrent Components. LSTM Unit store values for long time or for short time. LSTM has no activation functions for their recurrent components. Since there are no activation function the values ofunits does not change for some period until the context is changed. A LSTM Block contains such many units. LSTM's are considered as deep neural networks. These LSTM's are implemented in parallelsystems

iii. Encoders andDecoders

Encoder-Decoder LSTM approach has proven very

effective for seq2seq prediction problems. The architecture comprises of two models-one which is responsible for reading the input sequence and getting it encoded into a fixed-length vector and second one for decoding the fixed-length vector and to output the predicted sequence.

iv. Word Embedding

A word embedding is a learned portrayal for text where words that have the same meaning have a similar representation .Each word is mapped to one vector and the vector values are learned in such a way that represents a neural network, and thus thetechnique is frequently lumped into the field of deeplearning.

C. APPLICATIONS

The project is of use when there is a massive amount of data to be analyzed in order to draw a intuitive conclusion. Data nowadays is increasing at an exponential rate. Thereby giving rise to a problem of analyzing it and drawing various profitable results from it. This is where summarization comes into picture. Basically summarization is a technique in which large dataset is summarized into small countable lines of data, so that a user could get something meaningful out of it in a short span oftime. Thismodel also can help in reviewing websites, blogs, news articles, webpage and books. Data nowadays is generated from a lot of source as mentioned above. There are a lot of categories in which data may be segregated but segregating such a massive volume of data is not imaginable. So the model aims to generate summaries of the reviews provided on an e-commerce site in order to aid the firm in the decision making process to make insightful decision regarding thebusiness.

II. LITERATUREREVIEW

After surveying a number of resources and papers, a number of principles were understood and a gap analysis has been represented using a tabular representation in Table 1

TABLE 1
TABULAR REPRESENTATION OF THE VARIOUS LITERATURE SURVEY

TITLE	**AUTHOR AND YEAR**	**JOURNAL**	**STUDY/RESULT**	**GAPS IDENTIFIED**
Text Summarization Using Neural Networks	Khosrow Kaikhan,2004, [1]	IEEE	Training of neural network to learnthe relevant characteristics of sentences that should be included in the summary of the article.	If there is contradiction in data set, accuracy get reduced.
Machine Learning Approach for Automatic Text Summarization Using Neural Networks	Patel, Adwaita Chokshi, Satyadev Vyas, Khushbu Maurya,2018, [3]	IJARCEE	Address all the approaches to text summarization and present the modus operandi using machine learning	Didn't support diverse sentence simplification techniques and supporting executing it for multi-lingual and multiple

A Review Paper on Tex	NimishaDheer,Chetan Kumar,2016[7]	International Journal of Science and Research (IJSR)	describes the functionality of a comprehensive multidocument	documents conjointly.It also didn't provide essential elaboration.
Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond	Ramesh Nallapati, Bowen Zhou, Cicero dos Santos,2016,[4]	The SIGNLL Conference on Computational Natural Language Learning(CoNLL)	Modelling of abstractive text summarization using Attentional Encoder-Decoder Recurrent NeuralNetworks	Summary does not include important points
A Review Paper on Text Summarization	Deepali K. Gaikwad, C. Namrata Mahender,2016, [2]	IJARCCE	Provides an abstract view of the present scenario of text summarization	—
MEAD - a platform for multidocument multilingual text summarization	Dragomir Radev, Timothy Allison, Sasha Blair-Goldensohn[6]	IEEE	describes the functionality of a comprehensive multidocument multilingual summarization environment	Support for some languages are still unsupported and very hard to implement
Networks				Techniques can be refined and more features can be accounted for
Automatic Text Summarization: A Detailed Study	NimishaDheer,Chetan Kumar,2016[7]	International Journal of Science and Research (IJSR)	Studies the concepts,trends in text summarization and shows the importance of thestudy	Summary doesn't provide suffice articulation

III. PROBLEMDEFINITION

The problem definition in our project involves extracting key information from information source by employing a text summarization tool. This project aims to read the input file which may be in sort of documents to be analyzed and preprocessed so as to separate it into several segments that would each be attributed with values. These are then transformed into concrete semantic statements using several algorithms which can be mentioned. Our problem definition covers the concept of converting convoluted, mammoth data into a useful segment which is concise, yet doesn't lose the essence of original document.

IV. PROPOSED METHODOLOGY

Neural sequence-to-sequence models deliver a viable innovatory approach for abstractive text summarization (they are not limited to merely selection and rearrangement of passages fromtheoriginal text). However, these models have two shortcomings firstly they are prone to reproduction of inaccurate factual detail and moreover they hold a high tendency to repeat themselves [1]. In this work standard Long Short-Term Memory(LSTM) sequence-to- sequence attentional model has been used. This method uses a local attention model for generating each word of the summary reliant on the input sentence. While the model is structurally simple, it can easily be trained end- to-end and scalable to a huge amount of training data. The reconstructed paragraph was evaluated using standard metrics such as ROUGE, depicting that neural models are capable of encoding texts in sucha way that syntactic, semantic, and discourse coherence is preserved.The first step is to acquire data.The downloaded dataset,which is in a CSV file is read by a program and in second stepis cleaned.The preview is obtained. The cleaning phase consists of removing stopwords, removing null entries, replacing contractions with their longer forms, and removing unwanted characters like symbols and emojis. Stopwords are only removed from Text part of the reviews, they are not removed from Summaries to sound like more natural phrases. After processing the data, it will have only two columns i.e. text and summary, Once the data is cleaned, it is tokenized and converted into vectors so that it can be processed by the model.Then loading of the pre-trained ConceptNetNumberbatch word embedding takes place. The number of missing words in the embedding is found by comparing all the tokens from our reviews data to loaded embedding.Tokens such as <unk> are added for those words which are unknown and <pad> is a used in the end of line to indicate that it's the end of line. Finally, the sorting of the summary and text by the length of text, i.e. from shortest to longest is carried out. A Model will be designed which will be used while training. Since we are using TensorFlow,our major effort goes into building graphs. We will create few placeholders to hold our data and other parameters. Then we will create our bi-directional encoder, which will generate its own representation of inputdata.

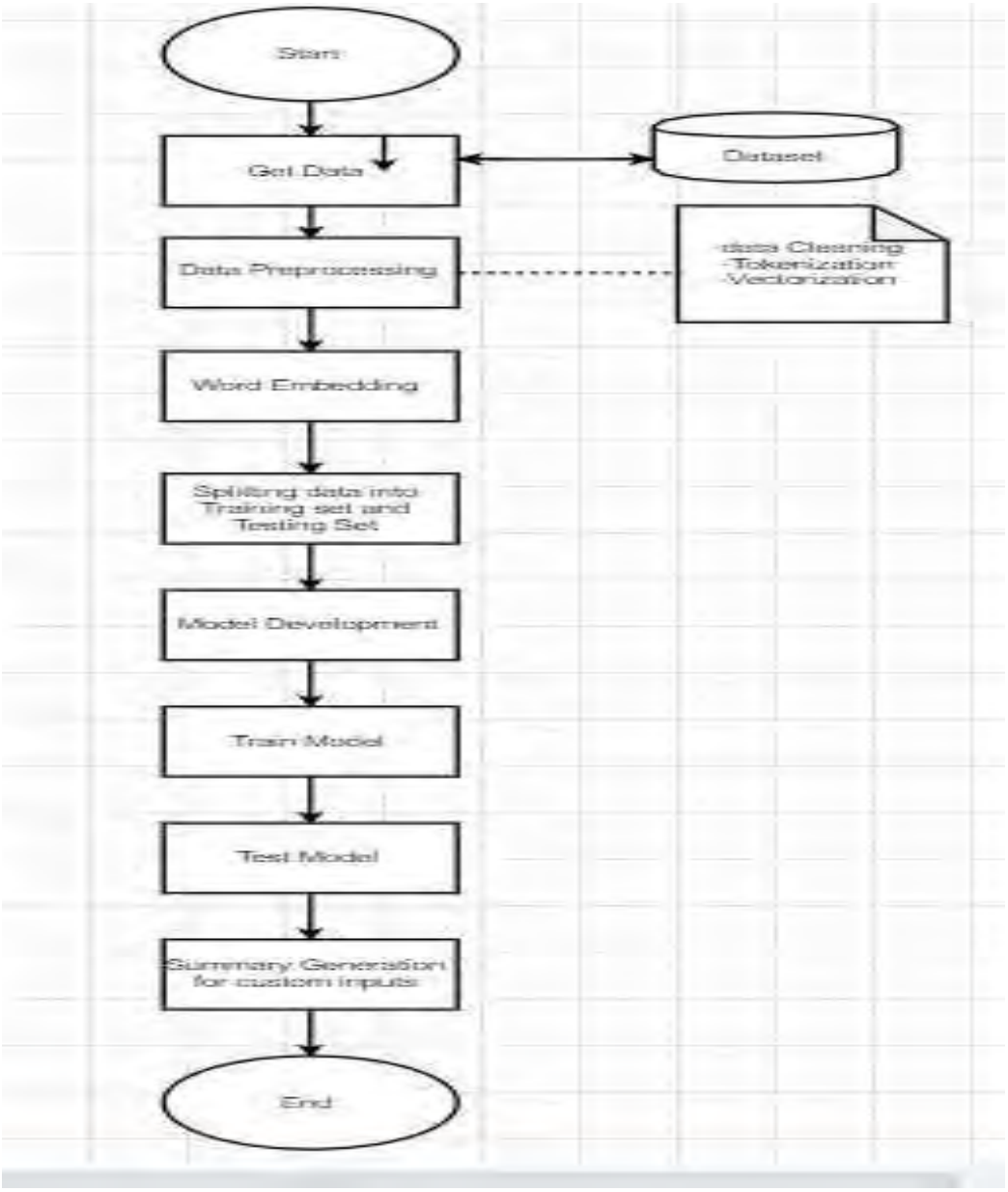


FIG 1:FLOWCHART DEPICTING SUMMARIZATION

V. RESULTS &DISCUSSIONS

The snapshots numbered Fig 2 and Fig 3 depict the initial preview of dataset and the results obtained post its cleaning. Thecleaningphase involves of removal of stopwords, null, null entries ,unwanted characters replacement of contractions with their longer forms, and removal of. After processing the data, it will have only two columns i.e. text and summary.

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1.0	1.0	5.0	1.303862e+09	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0.0	0.0	1.0	1.346976e+09	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJXXAIN	Natalia Corres "Natalia Corres"	1.0	1.0	4.0	1.219018e+09	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3.0	3.0	2.0	1.307923e+09	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0.0	0.0	5.0	1.350778e+09	Great taffy	Great taffy at a great price. There was a wid...

FIG 2: PREVIEW OF THE DATASET

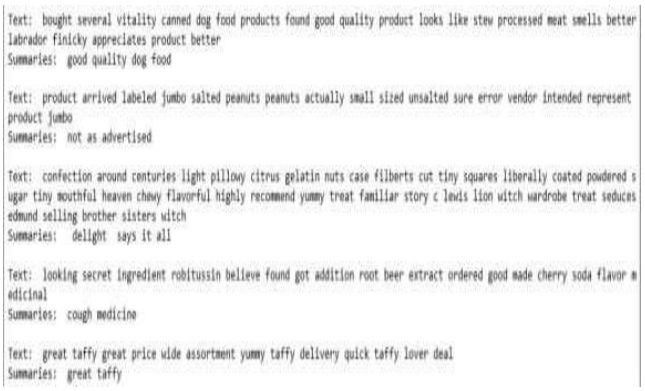


FIG 3: AFTER CLEANING THE DATA

Text Summarization is one of the inspiring field of research because it has many applications where we need summary of the large documents such as news headlines, market review, short messages on mobile, medical, business analysis etc.Abstractive summary methods produces highly relevant, crisp, information rich and less redundant summary, but because of its complexity less research has been carried out in the field of abstractive summarization. Although the interest in summarization has been developing and new approaches are being devised constantly, the task of automatic summarization has still unansweredquestions,
e.g. how to accomplish a deeper understanding of the document topic (natural language understanding, NLU), how to handle long documents, and how to enhance the evaluation strategies. Finding solutions to these questions will pave the way for further research in this area.

VI. ACKNOWLEDGEMENT

We sincerely thank the project co-ordinators for arranging the necessary facilities to carry out theproject work. We also extend our heartfelt thanks to our colleagues, family members and well- wishers.

References

[1]R. Brandow, K. Mitze and L. Rau,"Automatic condensation of electronic publications by sentence selection", Information Processing and Management, vol. 31(5), pp. 675-685,1995.

[2] P.B. Baxendale, "Machine-Made Index for Technical Literature: An Experiment" IBM Journal of Research and Development, vol. 2(4), pp. 354-361,1958.

[3] J. Kupiec, J. Pederson and F. Chen, "A Trainable Document Summarizer", Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, pp. 68-73,1995.

[4] M. Porter, "An algorithm for suffix stripping", Program, vol. 14(3), pp. 130-137, 1980.

[5] H.P. Luhn, "The Automatic Creation of Literature Abstracts", IBM Journal for Research and Development, vol. 2(2), pp. 159-165,1958.

[6] M.R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems", Journal of Research of the National Bureau of Standards, vol. 49, pp. 409-436,1952.

[7] I. Mani, Automatic Summarization, John Benjamins Publishing Company, pp. 129-165, 2001.

[8] W.T. Chuang and J. Yang, "Extracting sentence segments for text summarization: a machine learning approach", Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, pp. 152- 159,2000.

[9] C-Y. Liu and E. Hovy, "Identifying Topics by Position", Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP- 97), Seattle, Washington, pp. 283-290,1997

[10] AditiKonge, Manali Sarkar, RashmiHatwar, Vrushali Jain "Automatic Recapitulation of Text Document" in the Journal of IRJET published in Volume: 05 Issue: 03 inMarch-2018

[11] Deepali K. Gaikwad and C. NamrataMahender "A Review Paper on Text Summarization" in the journal IJARCCE published in Volume: 05 Issue: 03 inMarch-2016

[12] Reeve Lawrence H., Han Hyoil, NagoriSayaV., Yang Jonathan C., Schwimmer Tamara A., Brooks Ari D., "Concept Frequency Distribution in Biomedical Text Summarization", ACM 15th Conference on Information and

Deep Learning for Visual Question Answering

Vedant Maheshwari , Ms.Ashwini Patil , Tejas Garhewal , Luvnish Singh
Thakur College of Engineering Technology
Mumbai, India
vm6006@gmail.com , Ashwini.patil@thakureducation.org , scienticman@gmail.com , luvnish1998@gmail.com

Abstract— **Problems at the intersection of language and vision are of notable importance both for the diverse set of applications they offer and as challenging research questions. Introducing the task of open ended and free-form Visual Question Answering (VQA). A VQA system produces a natural language answer as the output while taking an image and a free-form, open-ended, natural language question about the image as inputs. Visual questions selectively target various areas of an image, including background details and underlying context. A system that succeeds at VQA typically needs a more comprehensive understanding of the image and complex reasoning than a system producing generic image captions. VQA is also amenable to automatic evaluation, since many open-ended answers contain only a few words or a closed set of answers that can be provided in a multiple-choice format. Visual questions target different areas of an image, including background details and underlying context.**

Keywords— Neural Networks, Text, Summarization, NLP, Deep Learning, Encoder Decoder

I. INTRODUCTION

Vision and language problems like visual question answering and image captioning have become more popular in recent years because the computer vision research community is progressing beyond recognition and towards solving multi-modal problems. The complex structure of language makes problems at the intersection of vision and language very challenging. We shall introduce the task of free-form and open ended Visual Question Answering (VQA). A VQA system takes as input a picture and a free-form, open-ended, tongue question about the image and produces a tongue answer because the output. This task which is goal-driven is applicable to scenarios like when visually-impaired users or intelligence analysts actively elicit visual information.
For our methods, we develop a 2-channel image (vision) question (language) model that ends with a softmax over K possible outputs. We choose the highest K = 1000 most frequent answers as possible outputs. The set of answers covers 83% of the train+val answers. This combined image + question embedding is then passed to an MLP– a fully connected neural network classifier with 2 hidden layers and 1000 hidden units (dropout 0.5) in each layer with tanh nonlinearity, followed by a softmax layer to obtain a distribution over K answers. The complete model is learned end-to-end with a cross-entropy loss. VGGNet parameters are frozen to those learned for ImageNet classification and not fine-tuned within the image channel. Although VGG is old, it was used because of its portability and ease of use, as well as due to the fact that it is very well understood.

II. HISTORY

A. VQA

Several recent papers have begun to review visual question answering. However, unlike our work, these are fairly restricted (sometimes synthetic) settings with small datasets. As an example, [1] only consider questions whose answers come from a predefined closed world of 16 basic colors or 894 object categories. [1] Also considers questions generated from templates from a hard and fast vocabulary of objects, attributes, relationships between objects, etc. In contrast, our proposed task involves open-ended, free-form questions and answers provided by humans. Our goal is to extend the range of data and sorts of reasoning needed to supply correct answers. Critical to achieving success on this harder and unconstrained task, our VQA dataset is 2 orders of magnitude larger than [1], [2] (>250,000 vs. 2,591 and 1,449 images respectively). The proposed VQA task has connections to other related work: Some has studied joint parsing of videos and corresponding text to answer queries on two datasets containing 15 video clips each. [3] Uses crowdsourced workers to answer questions on visual content asked by visually-impaired users. In concurrent work, [2] proposed combining an LSTM for the question with a CNN for the image to get a solution. In their model, the LSTM question representation is conditioned on the CNN image features at whenever step, and therefore the final LSTM hidden state is employed to sequentially decode the solution phrase. In contrast, the model developed during this paper explores “late fusion” – i.e., the LSTM question representation and therefore the CNN image features are computed independently, fused via an element-wise multiplication, then skilled fully connected layers to get a softmax distribution over output answer classes. It generates abstract scenes to capture visual sense relevant to answering (purely textual) fill-inthe-blank and visual paraphrasing questions. Some use visual information to assess the plausibility of sense assertions. One introduced a dataset of 10k images and prompted captions that describe specific aspects of a scene (e.g., individual objects, what is going to happen next). Concurrent with our work, some collected questions & answers in Chinese for COCO images. One automatically generated four sorts of questions using COCO captions.

B. Various Works

Text-based Q&A may be a well-studied problem within the NLP and text processing communities. Other related textual tasks include sentence completion. These approaches provide inspiration for VQA techniques. One key concern in text is that the grounding of questions. As an example, one synthesized textual descriptions and QA-pairs grounded during a simulation of actors and objects during a fixed set of locations.

VQA is of course grounded in images – requiring the understanding of both text (questions) and vision (images). Our questions are generated by humans, making the necessity for common- sense knowledge and sophisticated reasoning more essential.

Describing Visual Content- associated with VQA are the tasks of image tagging, image captioning, and video captioning, where words or sentences are generated to explain visual content. While the tasks require both visual and semantic knowledge, captions can often be non-specific. The questions in VQA require detailed specific information about the image that generic image captions are of little use [3].

Several papers have explored tasks at the intersection of vision and language that are easier to gauge than image captioning, like co reference resolution or generating referring expressions for a specific object in a picture that might allow a person's to spot which object is being mentioned. While concrete and task-driven, a limited set of visual concepts (e.g., color, location) tend to be captured by referring expressions. We shall demonstrate, a richer sort of visual concepts emerge from visual questions and their answers.

C. Problem Definition

The primary problem of our project is generating answers to visual questions which can be done using concepts of Image Recognition and Deep Learning. There are many potential applications for VQA. The foremost immediate is as an aid to blind and visually impaired individuals, enabling them to urge information about images both on the online and within the world. For instance, as a blind user scrolls through their social media feed, a captioning system can describe the image then the user could use VQA to question the image to urge more insight about the scene. More generally, VQA might be wont to improve human-computer interaction as a natural thanks to query visual content. A VQA system also can be used for image retrieval, without using image meta-data or tags. For instance, to seek out all images taken during a rainy setting, we will simply ask 'Is it raining?' to all or any images within the dataset. Beyond applications, VQA is a crucial basic research problem. Because an honest VQA system must be ready to solve many computer vision problems, it are often considered a component of a Turing Test for image understanding.

The overarching goal of VQA is to extract question-relevant semantic information from the pictures , which ranges from the detection of minute details to the inference of abstract scene 2 Figure 1: Object detection, semantic segmentation, and image captioning compared to VQA. The center figure shows the perfect output of a typical object detection system, and therefore the right figure shows the semantic segmentation map from the COCO dataset [10]. Both tasks lack the power to supply contextual information about the objects. The captions for this COCO image range from very generic descriptions of the scene, e.g., a busy town sidewalk next to street parking and intersections, too much focused discussion of one activity without qualifying the general scene.

D. Proposed Aspects of Project

This report is built to generate implementation approaches to address text summarization. This project will be helpful to the social networking or ecommerce companies, as well as to the blind people who want to get acquainted with technology. It will be helpful to these companies as these questions could then collect feedback and connections, and replies to various customer questions could then directly be generated. It shall benefit the blind man as solutions shall be given to their various questions, thus reducing human help and making them more capable.

This uses a classic CNN-LSTM model, where Image features and language features are computed separately and combined together and a multi- layer perceptron is trained on the combined features. While VGG Net is not the best CNN model for image features, GoogLeNet and ResNet have superior classification scores, but VGG Net is very versatile, simple, relatively small and more importantly portable to use. It takes as input a sequence of words and generates an output sequence of words. Extracting image features involves, taking a raw image, and running it through the model, until we reach the last layer. In this case our model is not 100% same as VGG Net, because we are not going to use the last two layer of the VGG. It is because the last layer of VGG Net is a 1000 way softmax and the second last layer is the Dropout. The question has to be converted into some form of word embeddings. Most popular software right now is Word2Vec. VQA is a simple model which combines features from Image and Word Embeddings and runs a multiple layer perceptron. By implementing this project not only are we enabling companies to save precious time and money, but also empowering various people by making it easier for them to join the technology ecosystem and live with it. Since the project works on images, combining them with the questions asked allows us to produce very accurate answers, with further proof that once enough data is provided to this model, it will produce even better results.

III. SYSTEM ANALYSIS

A. Methods

For our methods, we develop a two-channel image question model that culminates with a softmax over K possible outputs. We choose the top K = 1000 most frequent answers as possible outputs. This set of answers covers 82.67% of the train+val answers. We describe the different components of our model below.

Image Channel: This channel provides an embedding for the image. We experiment with two embeddings –

1. The activations from the last hidden layer of VGGNet are used as 4096-dim image embedding.
2. These are 2 normalized activations from the last hidden layer of VGGNet.

Question Channel**:** This channel provides an embedding for the question. We experiment with these embeddings:

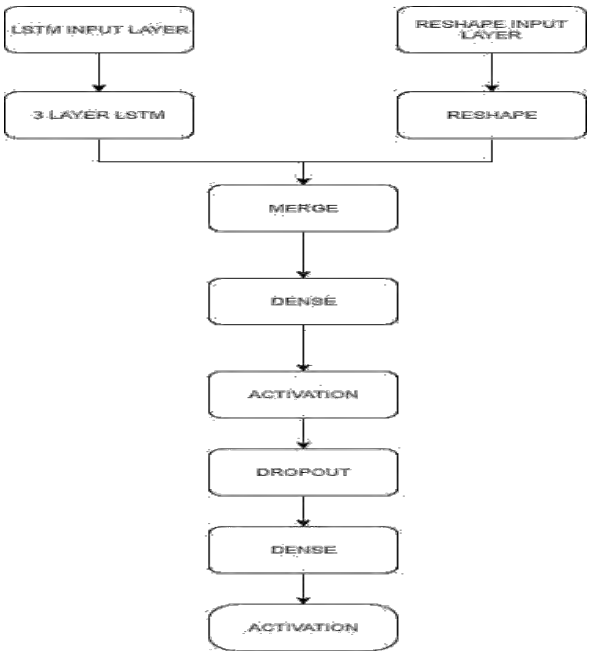


Figure 1

1)Bag of Words Question: the highest 1,000 words within the questions are used to create a bag-of-words representation. Since there's a robust correlation between the words that start a question and therefore the answer (see Fig. 5), we discover the highest 10 first, second, and third words of the questions and make a 30 dimensional bag-of- words representation. These features are concatenated to urge a 1,030-dim embedding for the question.

2) LSTM Q: An LSTM with one hidden layer is employed to get 1024-dim embedding for the question. The embedding obtained from the LSTM may be a concatenation of last cell state and last hidden state representations (each being 512-dim) from the hidden layer of the LSTM. Each question word is encoded with 300-dim embedding by a fully-connected layer + tanh non-linearity which is then fed to the LSTM. The input vocabulary to the embedding layer consists of all the question words seen within the training dataset.

3) Deeper LSTM Q: An LSTM with two hidden layers is employed to get 2048-dim embedding for the question. The embedding obtained from the LSTM may be a concatenation of last cell state and last hidden state representations (each being 512-dim) from each of the 2 hidden layers of the LSTM. Hence 2 (hidden layers) x 2 (cell state and hidden state) x 512 (dimensionality of every of the cell states, also as hidden states). this is often followed by a fully-connected layer + tanh non-linearity to rework 2048-dim embedding to 1024-dim. The question words are encoded within the same way as in LSTM Q.

Multi-Layer Perceptron (MLP): The image and question embeddings are combined to get one embedding. 1) For BoW Q + I method, we simply concatenate the BoW Q and that i embeddings. 2) For LSTM Q + I, and deeper LSTM Q
+ norm I methods, the image embedding is first transformed to 1024-dim by a fully-connected layer + tanh non-linearity to match the LSTM embedding of the question. The transformed image and LSTM embeddings are then fused via element-wise multiplication. This combined image + question embedding is then passed to an MLP – a totally connected neural network classifier with 2 hidden layers and 1000 hidden units (dropout 0.5) in each layer with tanh non-linearity, followed by a softmax layer to get a distribution over K answers. The whole model is learned end-to-end with a cross-entropy loss. VGGNet parameters are frozen to those learned for ImageNet classification and not fine-tuned within the image channel. We also experimented with providing captions as input to our model. We assume that a human-generated caption is given as input. We use a bag-of-words representation containing the 1,000 hottest words within the captions because the caption embedding (Caption). For BoW Question + Caption (BoW Q + C) method, we simply concatenate the BoW Q and C embeddings.

B. Technology Involved

There are primarily these approaches for achieving better troubleshooting results using the proposed technique: The question has got to be converted into some sort of word embeddings. Preferred is Word2Vec whereas lately state of the art uses skip-thought vectors or positional encodings. Extracting image features consists of taking a raw image, and running it through the model, until we reach the last layer. VQA may be a simple model which mixes features from Image and Word Embeddings and runs a multiple layer perceptron. Several technologies are used which include Keras+tensorflow framework, alongside training on VGGnet and Spacy. Keras is an open source,NN library written in Python. it's capable of running on top of TensorFlow, Theano, Microsoft Cognitive Toolkit or PlaidML.It was designed to enable fast experimentation with deep neural networks, it focuses on being extensible, modular and user-friendly. spaCy is an open-source software library for advanced NLP, written within the programming language Python. TensorFlow is an open-source software library for dataflow programming across a variety of tasks. it's a symbolic math library. VGG16 is a convolutional neural network model proposed by A. Zisserman and K. Simonyan from the University of Oxford within the paper "Very Deep Convolutional Networks for Large- Scale Image Recognition". Scikit-learn may be a free software machine learning library for the Python programing language. Scikit features various regression, classification and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is meant to interoperate with the Python's numerical and scientific libraries NumPy and SciPy.

C. User Interface Design

The UI has been kept simple and easily understandable due to the assumption that the target audience is not technology friendly. There is an upload button clicking on which will allow the user to upload the image. The box on the left of it shall show the file name and the text input box below it is for the user to explain the problem further. The troubleshoot button then submits this data further.

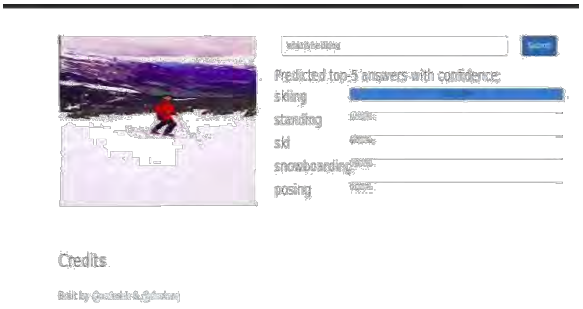


Figure 2

IV. CONCLUSION

As was expected, the vision-alone model that completely ignores the question performs rather poorly (open-ended: 28.13% / multiple-choice: 30.53%). In fact, on open-ended task, the vision-alone model (I) performs worse than the prior ("yes") baseline, which ignores both the image and question (responding to each question with a "yes"). To one's surprise, the language-alone methods which ignore the image perform surprisingly well, with BoW Q achieving 48.09% on open- ended (53.68% on multiple-choice) and LSTM Q achieving 48.76% on open-ended (54.75% on multiple-choice); both outperforming the closest neighbor baseline (open-ended: 42.70%, multiple- 10 choice: 48.49%). Our quantitative results and analyses suggest that this could be thanks to the language-model exploiting subtle statistical priors about the question types (e.g. "What color is that the banana?" are often answered with "yellow" without watching the image). For an in depth discussion of the subtle biases within the questions, please see. The accuracy of our greatest model (deeper LSTM Q + norm I selected using VQA test-dev accuracies) on VQA test standard is 58.16% (open-ended) / 63.09% (multiple-choice). We will see that our model is in a position to significantly outperform both the vision-alone and language-alone baselines. As a common trend, results on multiple-choice are better than open-ended. All methods are significantly worse than human performance. For some application domains, it might be useful to gather task-specific questions. As an example, questions could also be gathered from subjects who are visually impaired, or the questions could specialize in one specific domain (say sports). Bigham et al created an application that permits the visually impaired to capture images and ask open-ended questions that are answered by human subjects. These questions can rarely be answered using generic captions. Training on task-specific datasets may help enable practical VQA applications. We believe VQA has the distinctive advantage of pushing the frontiers on "AI-complete" problems, while being amenable to automatic evaluation. Given the recent progress within the community, we believe the time is ripe to require on such an attempt.

V. ACKNOWLEDGMENT

We sincerely thank our guide Ms. Ashwini Patil for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

References

[1] D. Geman, S. Geman, N. Hallonquist, and L. Younes. A Visual Turing Test for Computer Vision Systems. In PNAS, 2014.

[2] M. Malinowski and M. Fritz. A Multi-World Approach to Question Answering about Real-World Scenes based on Uncertain Input. In NIPS, 2014.

[3] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, and T. Yeh. VizWiz: Nearly Realtime Answers to Visual Questions. In User Interface Software and Technology, 2010.

[4] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase:A Collaboratively Created Graph Database for Structuring HumanKnowledge. InInternational Conference on Management of Data, 2008..

[5] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. H. Jr., and T. M.Mitchell. Toward an Architecture for Never-Ending Language Learning In AAAI, 2010.

[6] X. Chen, H. Fang, T. Lin, R. Vedantam, S. Gupta, P. Doll ar´, and C. L.Zitnick.Microsoft COCO captions: Data collection and evaluationserver.CoRR, abs/1504.00325, 2015.

[7] G. Coppersmith and E. Kelly. Dynamic wordclouds and vennclouds for exploratory data analysis. In ACL Workshop on Interactive Language Learning and Visualization, 2014.

[8] J. Deng, A. C. Berg, and L. Fei-Fei. Hierarchical Semantic Indexing for Large Scale Image Retrieval. In CVPR, 2011.

[9] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term Recurrent Convolutional Networks for Visual Recognition and Description. In CVPR, 2015.

[10] D. Elliott and F. Keller. Comparing Automatic Evaluation Measures for Image Description. In ACL, 2014.

[11] A. Fader, L. Zettlemoyer, and O. Etzioni. Paraphrase-Driven Learning for Open Question Answering. In ACL, 2013.

[12] A. Fader, L. Zettlemoyer, and O. Etzioni. Open Question Answering over Curated and Extracted Knowledge Bases. In International Conference on Knowledge Discovery and Data Mining, 2014.

[13] H. Gao, J. Mao, J. Zhou, Z. Huang, and A. Yuille. Are you talking to a machine? Dataset and methods for multilingual image question answering. In NIPS, 2015.

[14] D. Geman, S. Geman, N. Hallonquist, and L. Younes. A Visual Turing Test for Computer Vision Systems. In PNAS, 2014.

[15] J. Gordon and B. V. Durme. Reporting bias and knowledge extraction. In Proceedings of the 3rd Workshop on Knowledge Extraction, at CIKM 2013, 2013.

[16] S. Guadarrama, N. Krishnamoorthy, G. Malkarnenkar, S. Venugopalan, R. Mooney, T. Darrell, and K. Saenko. YouTube2Text: Recognizing and Describing Arbitrary Activities Using Semantic Hierarchies and ZeroShot Recognition. In ICCV, December 2013.

[17] M. Richardson, C. J. Burges, and E. Renshaw. MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text. In EMNLP, 2013.

[18] H. Agrawal, C. S. Mathialagan, Y. Goyal, N. Chavali, P. Banik, A. Mohapatra, A. Osman, and D. Batra. Cloudcv: Large-scale distributed computer vision as a cloud service. In Mobile Cloud Visual Media Computing, pages 265–290. Springer International Publishing, 2015.

[19] D. B. Lenat and R. V. Guha. Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project. Addison-Wesley Longman Publishing Co., Inc., 1989.

[20] M. Malinowski and M. Fritz. A Multi-World Approach to Question Answering about Real-World Scenes based on Uncertain Input. In NIPS, 2014.

Implementing System For Apparel Recommendation With Clothes Scouting Assistant Using Deep Learning

Shreyash Kad, Amey Pandya,Aishwarya Gupta, Shailesh Sangle
Computer Engineering Assistant Professor
Thakur College of Engineering
Mumbai, India
shreyashkad39@gmail.com, ameypandya@gmail.com, aishwaryagupta611@gmail.com, shailesh.sangle@thakureducation.org

Abstract— **Visual analysis of clothing is a topic that is receivingincreasedattentioninthefieldsofcomputervisionand machine learning. In our project, we determine what kind (genre) of clothing is worn in a set of pictures and using the involvement (likes) on these pictures to identify which genre is liked the most by the audience. After identifying the clothing in thepicture,thesystemhasafeaturetosearchforsimilartypeof apparelontheinternetandprovidetheuserwithlinks.System, also, provide a facility to user to find a person using the cloth and background descriptions. This model will be able to detect thegarmentitemsandclassifytheclothingattributesforfashion illustrations. This field has applications in e-commerce, social media and criminallaw.**

Keywords— Deep Learning, Computer Vision, Clothes Detection, Clothes scouting

I. INTRODUCTION (*HEADING1*)

ThisVisualanalysisofclothingisatopicthatisreceiving increased attention in the fields of computer vision and machine learning. Our project is based on the same idea. We aredevelopingasystemthataimstoidentifytheapparel worn by the user and the likes obtained on that post. Using the datasetprovidedtoit,itwillidentifywhatkindofapparel the user has worn. It will let the user know about the other users who have worn similar apparel and have obtained morelikes than them. Also, the system will also assist the user by providing some online shopping website links so as to assist the users to shop the same kind of apparel. This system will help the users especially the fashion bloggers who influence the fashion sense of theusers.

Besides this, the system will do comparative analysis using the likes obtained on the user's post and obtain other inputs from the posts posted by the other users. The working ofthesystemcanbeextendedfurthertoprovideacomparative basedmodelthatdoesacomparativestudyonvariousaspects with web scoutingassistance.

Inthelastfewyears,rapidprogresswasobservedindeep learning, especially in convolutional neural networks, GPU hardwareprocessingacceleration,andimageprocessing.With the help of multiple GPU cores, we can, now, train our deep learningmodelonmorethanamillionhigh-resolutionimages tolearnfeaturerepresentationandclassifyclothingimages.Afeature of our paper is to detect the proper landmark of the clothes from an image and to recognize its attributes and characteristics of that item. The same feature can be used to automatically tag the images with proper attributes of the clothes present in that image.

Since the purpose (or aim) of our system is to provide a holisticsolutiontotheinfluencersabouttheirclothing, wedo notjustleavetheinfluencerwithastatementabout"Whatare theylikedmorein"or"whatcreatesmoreengagementontheir socialmedia".Webelieveingoingastepaheadandproviding another feature wherein, after figuring out your most liked appareltype,theprojectwillscanforsimilarappareltypeson the web, which is up for sale, and provide the link (URL) of the same. This gives the clients direct access to clothes that theiraudiencelikestoseethemin,withanoptiontogoon and buythem.Oncetheattributesoftheclothesarecapturedfrom theinputimage,thoseattributescanbeusedtoscrapetheweb tofindsimilartypesofapparelavailableone-shopswiththeir prize.

Adding to the ground functionality of the project, our projecthasfeaturesonthesidewhichcanbeofadditional use in this same domain of image processing and clothing detection. The additional feature being referred to is to recognizeandtrackthemovementofthesuspect.Thesuspect can be identified by the clothes he or she is wearing and the fashion accessories (like a bag, eyewear, hat, etc.) he or she was having, even if the suspect's face is notvisible.

II. LITERATURESURVEY

A. A Deep-Learning-Based Fashion Attribute Detection Model

They propose a data-driven approach for recognizing fashion attributes. They engineered a modified version ofthe FasterR-CNNmodelwhichistrainedonalarge-scaledataset with 594 fine-grained attributes. They opted forDeepfashion as their Dataset for the category and attribute predictiontask. They removed 45 unclear attributes such as "Girl" "Please" andmergedsemanticallysimilarattributessuchas"geo","geo prints"and"abstractgeoprints".However,theywereno tableto deal with wrongly labeled images and cross-domain attributes.

Their architecture consists of a Faster R-CNN object detectionframeworkwithResNet101andROI-alignwithtwo modifications: a pruning mechanism and additional clothing attribute branches parallel to the category branch. They approach learning, both, attributes and category as a multi- task learningproblem.

B. Clothes Detection and classification UsingConvolution NueralNetworks

ThispapergaveaninsightintotheapplicationofCNNfor clothdetection.

They used:

a. CNNwith1ConvolutionalLayer

b. CNNwith3ConvolutionalLayer

c. CNNwith4ConvolutionalLayer

They had the following approach for evaluation:

I. Split the original training data into 80% training and 20% validation to optimize theclassifier.

II. Train the model for 10 epochs. They used categorical_crossentropylossfunctionandAdamoptimizer.

III. Theythentrainedthemodelandevaluatedthelossand accuracy. The following table depicts the loss and accuracy obtained using the above threemodels.

Table 1: Loss and Accuracy table

Number of CNN Layers	Test Scores	Test Accuracy
1	0.2484	0.9104
3	0.2496	0.9079
4	0.2656	0.952

C. Clothing Landmark Detection Using DeepNetworks With Prior of KeyPointAssociations

Thispaperconsiderstheproblemoflandmarkpointdetection inclothesfortheupperbody,whichisimportantandvaluable for the clothing industry. Their paper also discusses how to use these landmarks to recognize the category of clothing worn by the person in the image. The first step in their paper is to use YOLO object detection [7]. YOLO will form a boundingboxaroundthedetectedclothes.Thenextstep isto resizetheboundingboxto256x256,whichisusedasaninput tothefollowingstepofdetectinglandmark.Theyengineered adeepend-to-endarchitecturewhichhasfourbuildingparts:

1) A subnetwork of the stacked convolutional layer that captures the features of an inputlayer.
2) Adeconvolutionstackthatoutputsresolutionpreserving, each in the size of the inputimage.
3) Anotherlayerofconvolutionwhichisaninputtotheloss function.
4) Calculatingtheprobabilityofapointusing(2)and(3)to find the most possible candidates for thelandmark.

For identification of the attributes of the clothes,their algorithm focuses on a particular region. For example, the collar part will have four landmark points; theminimumrectangle that covers these landmark points is extended from width and height by 40%, respectively, with the original rectangle at the center position. Next, this rectangle is extractedfromtheimageandthefourlandmarksareconverted intoafeaturevectorasaninputtotheconvolutionsubnetwork basedVGG-13[4].Then,thefourfeaturevectorsareaveraged toonevectorasthefeaturerepresentationofthewholeimage. On top of this feature representation, we construct two fully connected layers, followed by three sibling vectors corresponding to the discriminative representation of the collar type, the sleeve type, and the clothing category, respectively. Finally, softmax loss functions are defined in these siblingvectors.
This paper, however, does not talk about the lower-body or other garment types apart from clothes for the upper-body.

D. Convolutional Neural Networks for Fashion Classification and Object Detection

They were able to achieve an accuracy of 50.2% for clothing style classification and clothing and 74.5% for attribute classification. They used Apparel Classification with Style (ACS) Dataset, which contains 89,484 images that were cropped based on bounding boxes around the individual's upper body. Each image is labeled with one of the 15 clothing categories.

For clothing type classification they used Histogram of Oriented Gradients (HOG), Speeded Up Robust Feature (SURF), LBP, and color information. Achieving35.03%, 38.29%, and 41.63% accuracy respectively. They then finely tuned R-CNN model to detect clothingobject

Multi-Label Fashion Image Classification with

Minimal Human Supervision

This paper properly deals with noisy data by building a baseCNNwhichisusedtoextractfeatures.ThebaseCNN is combined with a label cleaning network and an image classifier. The input images to their architecture were resized to 256x256. The base CNN observed about +7% increase.

III. SYSTEM MODEL

A. Dataset

1. DeepFashion

DeepFashion is a large scale clothes database, with over 800,000 diverse fashion images containing several appealing properties. The dataset is a label with fifty categories and is annotated with one thousand descriptive attributes. Out of total images, only 0.5% of images were found with inconsistent labels and required correctlabeling.

Theimagesinthedatasetaresegregatedintofivesub-datasetswhichareCategoryandAttributePrediction, Consumer-to-shopClothesRetrieval,In-shopClothes Retrieval, and LandmarkDetection.

1.1 Fashion LandmarkDetection:

The landmarks correspond to a set of key points defined in the structure of clothes, such as sleeves, cuffs, hemline, and neckline [2]. Landmarks are important because they will help in identifying the categories and attributes of the clothes. Landmark detection dataset contains 123,016 JPG images and“Anno” folder which contains pixel co-ordinates for bounding box and landmark coordinates. The main purpose of this dataset is used to train the Deep Learning model to detect. The Upper-body clothes possess six fashion landmarks, lower-body clothes possessfourfashionlandmarks,andfull-bodyclothes possess eight fashionlandmarks.

Table 2Fashion Landmark Dataset Label Classes used in Landmark Detection Task

Dataset Attributes	Classes
Clothes Type	Upper-body clothes, Lower-body clothes, Full-body clothes
Variation type	Normal pose, Medium pose, Large pose, Medium zoom-in, Large zoom-in
Landmark Visibility	Visible, Invisible/ Occluded, truncated/cut-off

1.2 Category and Attribute Prediction this dataset contains 289,222 JPG images and an “Anno” folder which containscoordinates

2. Street2Shop

Thegoalofthisdatasetistomatchtheapparelrecognized fromtherealworldimagetoasimilariteminanonlineshop. Thesedatasetaimstoshowcasehowonewouldstyleanoutfit andhelpothersdecidewhethertheywanttopurchaseanitem. Thedatasetcontainsthe“Photos”txtfileandJSONfiles. The “Photos” file contains URL to 424,840 images and supports 11 clothing articles. Supporting 11 clothing articles is one morereasontousethisdataset.Sothatourmodelwillals obe able to recognize more wearables other than upper body clothesandlowerbodyclothes.TherearethreetypeofJSON files, which are test_pairs_{category}.json, train_pairs_{category}.json and retrieval_{category}.json. Since Street2Shop Dataset supports 11 clothing categories. Therefore,thisdatasethas33JSONfiles.TheretrievalJSON filehasthelistoftheidofphotosthatcomeunderitscategory. The test and train JSON file have a list of photo IDs and the co-ordinatesoftheboundingbox.EachJSONfilecontains an image ID of an image, which belongs to that particular category, and product ID. The dataset contains two types of images:

1) Street Photos: There are large variations in theimages. Variations in terms of lighting, indoor vs outdoor environments, body shapes and sizes of the people wearing theclothing,poseofthepeople,cameraviewingangle,anda huge amount of occlusion due to the layering of items in outfits.Inaddition,aphotomaydepicteitherahead-to-toeThe dataset is available online at http://www.tamaraberg.com/street2shop. It contains two files a “photos.txt” file and a “meta.zip” file. “Photos.txt” file contains the id of the image and its URL separated by commas. We wrote a python program to download those images and save them using their ID as the name.

B. SystemSpecification

Like other deep learning problems, it is crucial to have a computational system with powerful GPUs to train a deep learning model on a hefty amount of data..

Table 3 System Specifications

Component	Specification
RAM	16 GB
CPU	Intel I7
GPU	
Cuda Toolkit	10.2
Operating system	Ubuntu 18.04

Table III gives the detail configuration of the laptop on which we have carried out the experiments. We installed CUDA, a parallel computing platform and programming model developed by NVIDIA [11], to utilize the power of GPU cores.

Every method of this paper is implemented in Pytorch as it enables easy, fast, flexible experimentation and efficient production[10].Pytorchiseasytoperformexperiments ,asits syntaxisverysimilartothatofpython,yetprovidespowerful APIs and ecosystem. One of the main reasons for opting for pytorch is its support for data parallelism. Because of Data parallelism,thepytorchcandistributethecomputationalwork among multiple GPU cores which results in reduced training and testingtime.

C. Proposed Method

1. Clothing item and AccessoriesDetection

Itisveryimportanttodetectnotonlyoneclothingitem from the given frame but all clothing and fashion accessories like bags, belts, hats and so on. In this paper, our goal is to detect upper-body, lower-body and/or both full-body apparels and accessories that includehats,eyewear,footwear,bags,andbelts,ifany ofthemarepresent.Forthispurpose,wefine-tunethe pre-trained Faster-RCNN model to detect the present objects [13]. We created two models, the first model was trained on the DeepFashion dataset (Landmark detection benchmark) and second onStreet2Fashionshotorseveralpartial-bodyshots.Thesevividvarietiesreflect the extreme challenges and variations that we expect to find for clothing retrieval in real-worldapplications.

2) ShopPhotos:Theseareimagesoffashionmodelsorjust clothes from an e-commerce site. This dataset has 404,683 imagesfrom25differentonlineclothingretailersandwith11 clothing articles starting from small items as belts and eyewear, to medium size wearables like hats or footwear, to larger items as jackets, tops, orpants.

Collection of Street2Shop dataset:Dataset. The need to train the model on the Street2Fashion dataset was for the model to identify thefashionaccessories.Beforetraining,allimagesare resized to 512 x 512 size and are passed through the fastNIMeansDenoisingColored method of OpenCV packagetodenoisetheimages[12].Themodeloutputs the image with multiple bounding boxes around the clothes and other fashion accessories. In addition to this, the details are stored in a JSON file, an example is shown in figure I so that it can be used for further experiments.

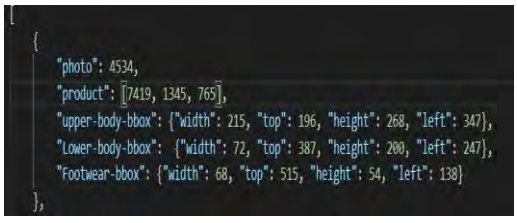


Figure IOutput JSON file of Clothes Detection Model

ThisoutputJSONfilecontainsthephotoid,productid and bounding box location for present clothing and accessories items. Table IV shows which color is assigned to which FashionItems.

Table 4 Color assigned to the bounding box

Items	Color assigned to the bounding box
Upper-Body Clothes	Green
Lower-Body Clothes	Yellow
Full-Body Clothes	Blue
Other Fashion accessories	Red

2. LandmarkDetection

Beforebeginningwiththelandmarkdetection,weuseour fine-tuned faster RCNN model to detect clothes and accessoriesandformboundingboxesaroundthesame. From the output JSON file, which is obtained from the detection process, we will focus only on bounding boxes for clothing (upper-body, lower-body, and full-body clothes). Then these selected regions of the image by bounding boxes arecropped from the image shown in figureII.

Figure II Cropping of the required region for landmark detection

Once the required regions are cropped out of the given imagestheyareresizedto512x512.Now,eachnewimageis passed through our custom made deep architecture, which is inspired by[3].

2.1. Deep Architecture

Figure III shows the deep architecture of our model and is drawn with the help of the online toolNN_SVGprovidedbyAlexNail[14].Inthis figure, the first two layers are the convolutional layersofsize512x512x64.Followedbyamax-poollayerof216x216x128andagainfollowed bytwoconvolutionallayers.Thisisrepeateduntil the size layer becomes 64 x 64. Since there are three max-pool layers, the size of the feature is too small to capture sufficient information to pinpointthelocationofthelandmark.Hence,we use three deconvolutional neural layers to map theoutputfeaturesonthesameimagesizethatof the input, which is 512 x 512. At the end of the architecture, the softmax function is used to

compute the probability for each landmark detected. For upper-body clothes, our network provide12 landmark points. For lower-body clothes,thenetworkprovides8landmarksand14 landmarks for full-bodyclothes.

3. Clothing attributeclassification

   To predict the category and attributes of the clothes presentinanimage,wewillusethelandmarkfeatures of the clothes extracted by the erstwhile network. We propose another network of neurons to recognize the attributesofapparelasindicatedinfigureIV.Indetail, the network uses a layer of dimension 2048 x N x N, where N is the number of landmarks provided by the previous layers as per the clothes location, three convolutionallayersfollowedbyalayerofninenodes, for each attribute except for the color, with softmax activationfunction.Forcoloridentification,thispaper usesfocusesontheregionwithintheboundingbox.If the same color is present in the maximum region then that color is declared as the color of the clothing, for which we trying to predict theattribute.

4. Clothes ScoutingAssistant:

   This feature can be implemented simply thanks to the vivid python web scraping packages and google's shopping search feature. The attributes predicted can beusedaskeywordstosearchtheshoppingsectionof google to find clothes that are similar to detected clothes.

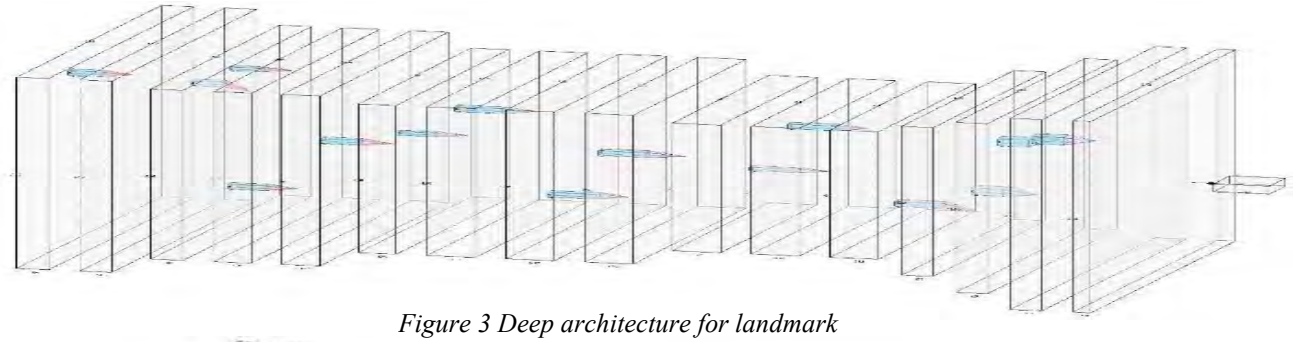


Figure 3 Deep architecture for landmark

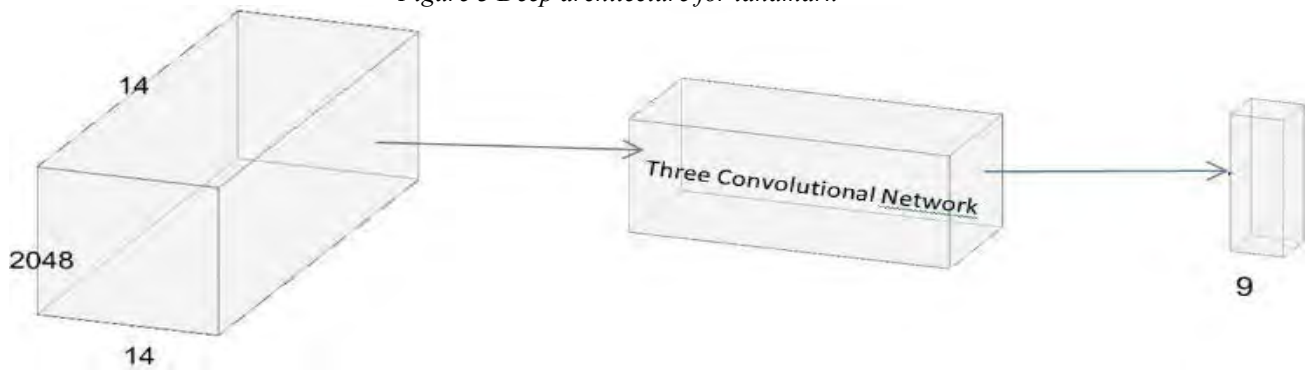


Figure 4 Clothes scouting output

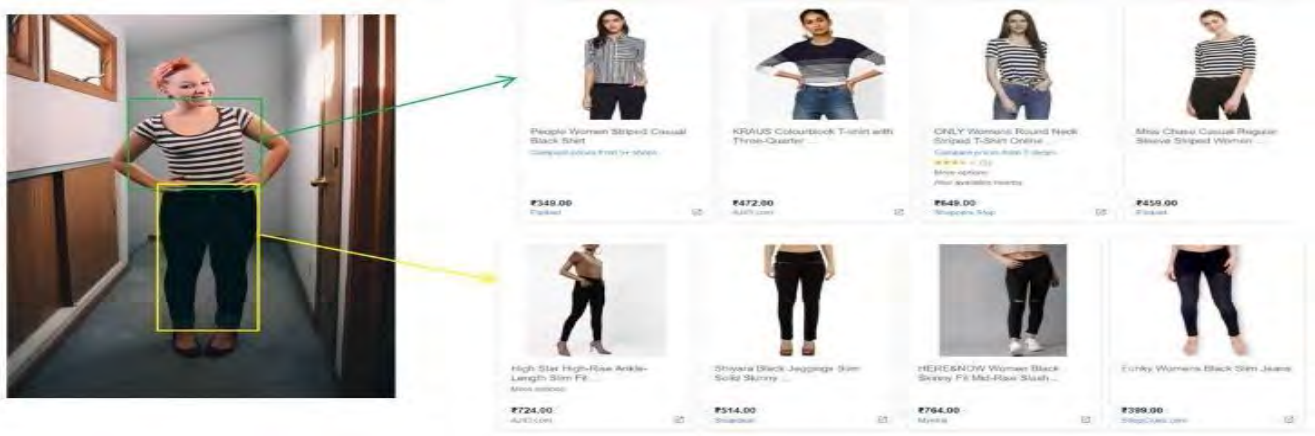


Figure 5 Clothes scouting output

IV. EXPERIMENTALRESULTS

A. *Clothing item and accessoriesdetection*

This model was trained on both datasets for 1000 epochs, that is, learning used no less than 90% of GPU's processing power. Still, there were small items like footwear or eyewear, in some images that the model was failed to detect. Figure VI (i& iii) depicts such examples where the footwear was not detected in (i) and the bag was not detected in (iii).

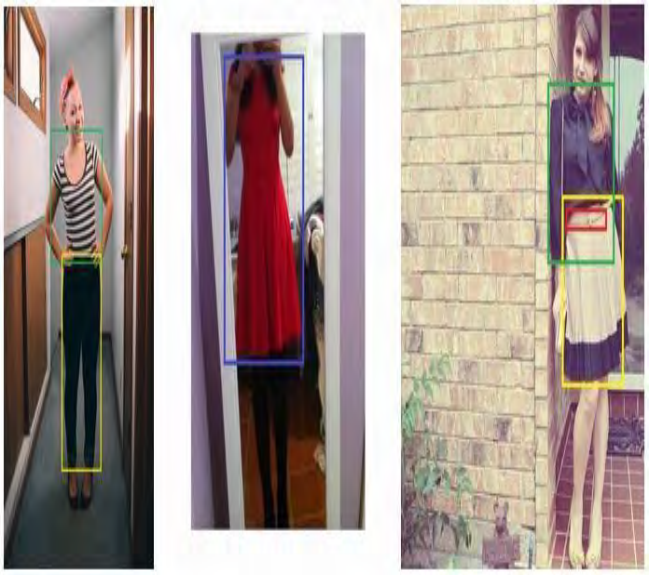


Figure 6 Clothes detection results

B. *LandmarkDetection*

To evaluate the Landmark detection model, we used normalized mean error as the evaluation metrics. The Thismodelwastrainedonbothdatasetsfor1000epochs ,thatnormalized mean error is defined as,

$$Normalized\ mean\ error\ = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{\left|L_{(j)}^{(i)} - L_{(j)}^{(i)}\right|}{k\ x\ D_i}$$

Where $L_{(j)}^{(i)}$ represents the coordinate of the ith landmark point in the jth image, $L^{(i)}$represents the ithlandmark

coordinates of the jth image, k is the total number of landmarks, N is the number of test images, and Di is a normalized factor which is defined as Euclidean distance between the landmark points of left-shoulder to right shoulder in the ith image.

In the experiment, we used the training, testing, and validation images provided by the DeepFashion (Landmark prediction benchmark) dataset. Our model scores a normalized mean error of 0.1743. Figure VII shows some outputs of the model.

Figure 7 Clothes detection results

C. *Clothing Attribute Classification*

Since we created two models and the first model was trained and validated on the DeepFashion dataset and the second on the street2Shop. Once, the validation result reached an accuracy of more than 80%, we started the testing phase. The first model was tested on the street2Fashiondatasetandthesecondmodelwastested on the DeepFashion one. The images with the wrong prediction were evaluated in person and if any error was found in labeling, it was corrected manually. In this way, we cleaned the annotations of both datasets to increase their accuracy. Again the models were trained but now withtheviceversadatasets.Weevaluatedourmodelusi ng three evaluation metrics: recall, precision, and F1-Score and the overall score was 91%. Figure VIII gives the distributed scores for each attributeclass.

Figure VII Clothing attribute classification.

V. CONCLUSION

In this paper, we engineered an intelligent system based on deep learning to find the landmarks of the detected clothes and to find the attributes of the apparel using the landmarks.This could be done because of the datasets DeepFashion and Street2Shop, which are openly available. The system also focuses on social media and provides

	precision	recall	f1-score	support
Class 0	0.82	0.91	0.86	1000
Class 1	0.99	0.99	0.99	1000
Class 2	0.82	0.90	0.86	1000
Class 3	0.95	0.92	0.93	1000
Class 4	0.85	0.90	0.87	1000
Class 5	0.99	0.97	0.98	1000
Class 6	0.84	0.64	0.73	1000
Class 7	0.94	0.98	0.96	1000
Class 8	0.98	0.98	0.98	1000
Class 9	0.97	0.96	0.97	1000
micro avg	0.91	0.91	0.91	10000
macro avg	0.91	0.91	0.91	10000
weighted avg	0.91	0.91	0.91	10000

an efficient and user- friendly system. Also by

using the system can be used to identifytheprobableplacewheretheuserwouldhaveclicked thepictureusingthebackgroundimage.Thisideologycanbe used in various fields like criminallaw

VI. References

[1] A Deep-Learning-Based Fashion Attributes DetectionModel

[2] J. Cychnerski, A. Brzeski, A. Boguszewski, M. Marmolowski, M. Trojanowicz, "Clothes Detection and Classification Using Convolutional Neural Networks", Proceedings of 22nd IEEE International Conference on Emerging Technologies and Factory Automation, pp. 1-8,2017

[3] ClothingLandmarkDetectionUsingDeepNetworksWithPriorofKey PointAssociations

[4] Convolutional Neural Networks for Fashion Classification and Object Detection

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014. [Online]. Available:https://arxiv.org/abs/1409.1556

[6] http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html

[7] https://liuziwei7.github.io/papers/deepfashion_supp.pdf

[8] M.HadiKiapour,XufengHan,SvetlanaLazebnik,AlexanderC.Berg, Tamara L. Berg. Where to Buy It: Matching Street Clothing Photos in Online Shops. International Conference on Computer Vision (ICCV), 2015.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Las Vegas, NV, USA, 2016, pp. 779– 788.

[10] https://pytorch.org/features/

[11] https://developer.nvidia.com/cuda-zone

[12] https://docs.opencv.org/3.4/d1/d79/group photo denoise.html#ga0 3aa4189fc3e31dafd638d90de335617

[13] R. Girshick. Fast R-CNN. In ICCV,2015

[14] http://alexlenail.me/NN-SVG/AlexNet.html

[15] Y.Yorozu,M.Hirano,K.Oka,andY.Tagawa,"Electronspectroscopy studiesonmagneto-opticalmediaandplasticsubstrateinterface,"IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301,1982].

Implementing Predictive Analysis System Of Medicines And Doctors Availability In Government Hosptials

Vikas Singh, Adit Rathi, Kevin Shah, Siddharth Tiwari, Mukul Suvarna
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
vikassngh@gmail.com, aditrathi98@gmail.com, kevinshah05@gmail.com, siddharthtiwari1212@gmail.com, mukulsuvarna15@gmail.com

Abstract— **The social issue of government hospitals i.e. availability of medicines and doctors is a troublesome one. Using the knowledge of software development, web development, analysis of the data and management of information systems, we can come up with an effective solution. This portal will enable the common people to identify best medicine for their symptoms and best doctor to treat that disease. It will provide the best suited medicine for the entered symptom. The interface will ask the user for various symptoms and duration of occurrence and then, the system will provide with necessary medicines and the seriousness of the symptoms, availability of doctors in the local hospitals.**

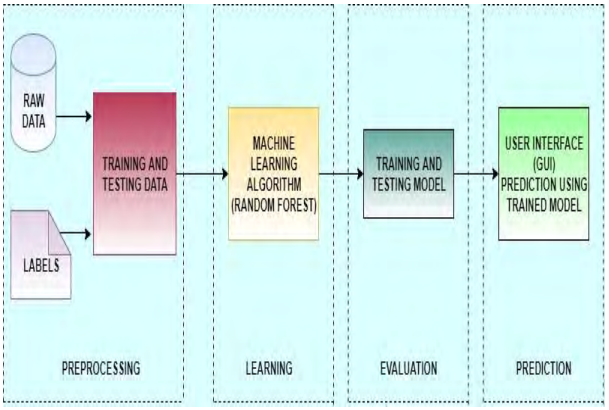
Keywords— Decision Tree, Naïve Bayes, Random Forest Regressor, Prediction, Data Extraction.

I. INTRODUCTION

In recent years, with the development of computer technologies, the puzzles with respect to Bayesian statistics and posterior distribution have been better solved. Meanwhile, Bayesian statistics successfully apply to economic, sociology and some other fields. In medical fields, the foreign scholars have solved some medical problems that are hard to be settled in classic statistics by using Bayesian classification. Naive Bayes is one of the most popular classification technique introduced by Reverend Thomas Bayes.
It is the application of computing and communication technologies to optimize health information processing by collection, storage, effective retrieval (in due time and place).The proposed system is mainly used by the all the people where confidentiality and integrity of the data has utmost importance. Computer assisted information retrieval may help support quality decision making and to avoid human error. Imagine a doctor who has to examine 5 patient records; he or she will go through them with ease. If the number of records grows with a time constraint, it is almost certain that the accuracy with which the doctor delivers the results will not be as high as the ones obtained when he had only five records to be analyzed.
At present in order to remain healthy, regular body diagnosis is necessary. Today, there are multiple sources available as individual prediction or recommendation system but the need of the hour is to have an integrated model comprising both. Also, it would be more appropriate and convenient if people could get basic diagnosis online 24x7 rather than visiting hospitals & clinics frequently. Thus, reducing cost and saving time. If certain anomalies found in the diagnosis then recommendation of nearby specialist and hospitals according to user's preference would facilitate in quick and appropriate treatment.

II. METHODOLOGY

The overall working of our project can be explained through the following abstract diagram. The high-level methodology used in this project is as follows:
Disease Prediction has been already implemented using different techniques like Neural Network, decision tree and Naïve Byes algorithm. Particularly heart related disease is mostly analyzed. From the analysis it was found that Naïve Bayes is more accurate than other techniques. So, Disease Predictor also uses Naïve Bayes for the prediction of different diseases.



The methodology consists of 3 main stages- Data Pre- processing, Data Preparation, Encoding the data.

A. Naïve Bayes

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. The dataset is divided into two parts, namely, feature matrix and the response vector.

- Feature matrix contains all the vectors(rows) of dataset in which each vector consists of the value of dependent features. In above dataset.
- Response vector contains the value of class variable (prediction or output) for each row of feature matrix.

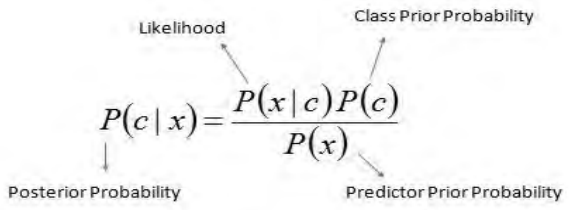
The fundamental Naive Bayes assumption is that each feature makes an:

- Independent
- Equal

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where A and B are events and P(B) = 0.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.

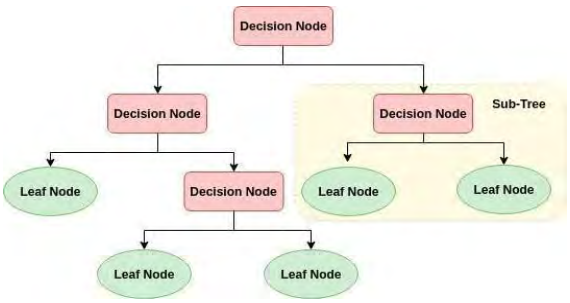
P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance (here, it is event B).

P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen.

Naive Bayes algorithms are mostly used in sentiment analysis, spam filtering, recommendation systems etc. They are fast and easy to implement but their biggest disadvantage is that the requirement of predictors to be independent. In most of the real-life cases, the predictors are dependent, this hinders the performance of the classifier.

B. Decision Tree Classifier

A decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret.



White-box models are the type of models which one can clearly explain how they behave, how they produce predictions and what the influencing variables are. There are two key elements that make a model white-box: features have to be understandable, and the ML process has to be transparent. Decision Tree is a white box type of ML algorithm. It shares internal decision-making logic, which is not available in the black box type of algorithms such as Neural Network. Its training time is faster compared to the neural network algorithm. The time complexity of decision trees is a function of the number of records and number of attributes in the given data. The decision tree is a distribution- free or non-parametric method, which does not depend upon probability distribution assumptions. Decision trees can handle high dimensional data with good accuracy.

We identified certain problems with SVM which is why we could not use it in the project. Some of them have been mentioned below.

1. A small change in the data can cause a large change in the structure of the decision tree causing instability.
2. For a Decision tree sometimes calculation can go far more complex compared to other algorithms.
3. Decision tree often involves higher time to train the model.
4. Decision tree training is relatively expensive as complexity and time taken is more.

C. Random Forest

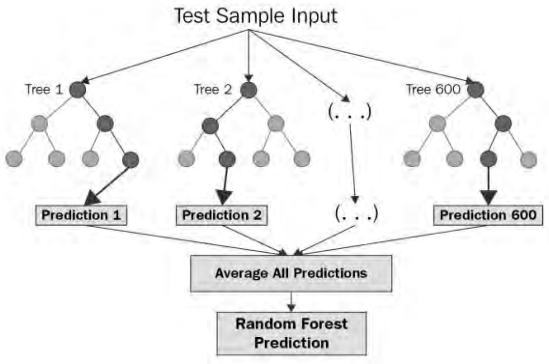
Random forests is a supervised learning algorithm. It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance.

Random forests has a variety of applications, such as recommendation engines, image classification and feature selection. It can be used to classify loyal loan applicants, identify fraudulent activity and predict diseases. Random forests is considered as a highly

accurate and robust method because of the number of decision trees participating in the process. It does not suffer from the overfitting problem. The main reason is that it takes the average of all the predictions, which cancels out the biases. The algorithm can be used in both classification and regression problems. Random forests can also handle missing values. There are two ways to handle these: using median values to replace continuous variables, and computing the proximity-weighted average of missing values. You can get the relative feature importance, which helps in selecting the most contributing features for the classifier.Random forest is a bagging technique and not a boosting technique. The trees in random forests are run in parallel. There is no interaction between these trees while building the trees. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

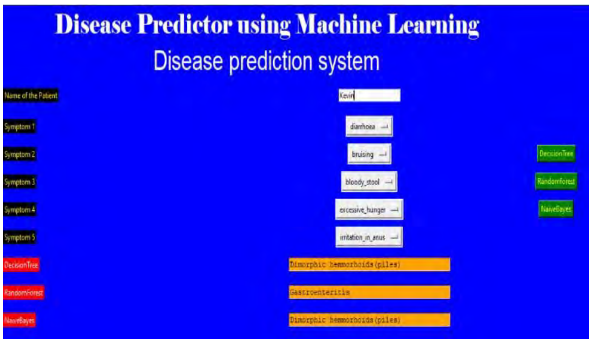
A random forest is a meta-estimator (i.e. it combines the result of multiple predictions) which aggregates many decision trees, with some helpful modifications:

1. The number of features that can be split on at each node is limited to some percentage of the total (which is known as the hyperparameter). This ensures that the ensemble model does not rely too heavily on any individual feature, and makes fair use of all potentially predictive features.
2. Each tree draws a random sample from the original data set when generating its splits, adding a further element of randomness that prevents overfitting.



III. PROPOSED WORK

An intelligent system for accurate disease prediction and medical facilities recommendation plays a major role in effective treatment. This system takes the symptoms from users and predicts the most accurate disease accordingly. Additionally, sensor module helps in continuous evaluation of vitals like heart rate, blood pressure and sugar level for patient which is fed in the system at runtime for analysis along with other external symptoms. Based on the prediction, system recommends the doctors according to user's preference out of nearest location with doctors having expertise for that particular disease to avail the required medications. Also, the users can provide their feedback for the recommended doctors.



IV. CONCLUSION

There are challenges to overcome, such as the high degree of variability and the degree of difficulty in complete capture and access; however, even with these challenges, the potential for improving patient care and hospital efficiencies is immense.

The number of medications a patient takes is predictive of medical complexity with simple models based on counts of the number of medications as effective as more complex models.

V. ACKNOWLEDGEMENT

We sincerely thank our guide Mr. Vikas Singh for his guidance and constant support. We also thank the project coordinators for arranging the necessary facilities to carry out the project work. We thank the HOD, Dr. Sheetal Rathi, Dean Academic Dr. R.R Sedamkar, the Principal Dr. B. K. Mishra and the college management for their support.

References

[1] Binal A. Thakkar, Mosin I. Hasan, Mansi A. Desai, "Healthcare decision support system for swine flu prediction using naïve bayes classifier", "IEEE", 101-105, 2010.

[2] Feixiang Huang, Shengyong Wang, Chien Chung Chan, "Predicting disease by using data mining based on healthcare information system", "IEEE", 2012. [3] Rafael M. D'addio, Marcelo G. Manzato, "A collaborative filtering approach based on users' reviews","IEEE", 204-209, 2014.

[4] F. O. Isinkaye, Y.O. Fola Jimi, B.A. Ojokoh," Recmmendation systems : Principles, Methods and Evaluation", "Elsevier", 261-273,2015.

[5] Manjusha K.K.,K. Sankaranarayanan, Seena P,"Prediction of different dermatological conditions using naïve Bayesian classification","IJARCSSE",864-868,2014 [6] Subhash C. Pandey, "Data Mining techniques for medical data: A Review", "IEEE", 2016.

[7] Lee, Yunkyoung, "Recommendation system using collaborative filtering", 2015.

[8] Geoff A., Elvis C., Mickey E., Sameul S., "Smart wearable body sensors for patients self assistant and monitoring", BioMedCentral, 2014.

[9] G Gannu, Y Kakodkar, A Marian, "Improving the quality of predictions using textual information in online users reviews", Information Systems, 2013.

[10] Sellappan P., Rafia A., "Intelligent heart disease prediction system using data mining techniques", IEEE/ACS, 2008.

Design Of System To Efficiently Choose Machine Learning Algorithms

Atharva Tendulkar,Athashree Vartak, Kasturi Dhotre, Mrs. Shiwani Gupta
Computer Engineering Departnment Thakur College of Engieerning and Technology, Mumbai.
tendulkaratharva@gmail.com,athashreeav5556@gmail.com,kasturid890@gmail.com,shiwani.gupta@thakureducation.org

Abstract— **In Machine Learning, the primary problem faced isn't implementation of algorithm but choosing the algorithm to be implemented for the chosen dataset. Thus, the common approach followed is implementing all the feasible algorithms and then choosing the one that performs best. The main drawback of this approach is that a hefty amount of time is wasted for training and testing data on myriad of algorithms. On the other hand, this time could be saved by understanding the mathematical approach used by the algorithms for fitting and evaluating the data to understand it's working. The approach followed here is based on the understanding of the algorithms and their background mathematical functioning to deduce the type of data it can perform well. The type of data here means the types of features (categorical or discrete), correlation among the independent variables (high or low) and the correlation of independent variables with dependent variables, etc.**

Thus, this approach will help in saving significant amount of time as a single or few numbers of algorithms are chosen instead of trying out many algorithms and then determining which works the best. But, here, the data will be needed to pre-processed and visualized to understand its nature in order to determine which algorithm is to be chosen. The visualization plays a crucial role as it helps in understanding the underlying information that cannot be realised in just a glance at the data and in turn requires series of processing and analysis.

Keywords: Machine Learning, Classification, Efficient, features, visualization

I. INTRODUCTION

Business practices and habits are very difficult to change especially in large and mature companies. Introducing new ways of working is extremely difficult, but making people work differently is almost impossible. There is huge cultural inertia in these companies and business industry in general. Use of Machine Learning is one of those changes that will make people work differently and will make business environments different in future.Machine Learning is a field of computer science which gives the computer the ability to deduce patterns and learn without being programmed explicitly. Machine Learning's primary aim is to provide training to an algorithm to perform required tasks. Machine Learning is one of the most important technologies with applications like Speech recognition, recommendations, fraud detection, and financial trading, etc. It contains various types like Supervised Learning, Unsupervised Learning, Semi- supervised Learning, and Reinforcement Learning which each has their own use cases and algorithms. Supervised Learning maps an input to an output based on example input-output tuples which is also known as labeled data. Machine Learning has 2 subclasses ie. classification and regression. A job is a classification task if the output is categorical and a regression task if the output is a continuous value.

II. PROPOSED WORK

The standard approach to standard machine learning models includes the following steps: Firstly, data needs to be acquired and pre-processed. Then, appropriate algorithm is applied to get the result. Then the result is evaluated on the basis of a variety of performance measures dependent upon the algorithm used.

The proposed method primarily includes detailed study of classification algorithms and understanding the mathematical basis of its functioning. This study will make the process of understanding and interpreting the results easy. The processing of determining the algorithm to be chosen for the required dataset requires pre-processing of data and visualization. Visualization plays a crucial role in understanding the data and its features.After detailed study of classification algorithms like Logistic Regression, Naïve Bayes, k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, etc appropriate datasets are to be acquired. The process of obtaining the datasets is tedious as all the classification algorithms listed above have a different set of requirements and limitations like a particular algorithm might only need categorical values or maybe some other algorithm might not function well for correlated data, etc. Thus, such kinds of assumptions require a dataset that should follow all the expected conditions or if possible, the dataset needs to be transformed into the one that can be used as per the requirement.

The transformation mentioned here can include dealing with missing values by replacing the values by mean, median or applying another prediction algorithm to determine the missing value. It may also include scaling the the attributes, dealing with imbalanced datasets, and so on. Thus, after finalizing datasets, the algorithms must be applied upon the dataset decided. After training and testing the model, evaluation needs to be carried out on the basis of performance measures. For example,. In case of

Logistic regression, F1 score and ROC are better performance measures than accuracy. Therefore, each and every algorithm is to be evaluated on the basis of the appropriate performance measures. After evaluating the models, they need to be optimized. The process of optimization will differ from algorithm to algorithm. Hyper-parameter tuning is one of the methods of optimization. For example, in KNN the optimum value of k can be found using elbow method to optimize the result. On the other hand, in case of decision trees, the tree can be pruned to obtain the optimum results.

A. Advantages

- Less Time
- Understanding how algorithms actually works would help to give optimized results and will aid in interpreting results more efficiently
- Knowledge of algorithms and its mathematical working will help in understanding parameter tuning.

B. Methodology

For this project, Scrum methodology is used which is an agile software development methodology. On scrum projects, every member of the project team has the opportunity to know how the project is going at any given time. A Scrum process is distinguished from other agile processes by specific concepts and practices, divided into the three categories of Roles, Artifacts, and Time Boxes. Scrum is most often used to manage complex software and product development, using iterative and incremental practices. Scrum significantly increases productivity and reduces time to benefits relative to classic "waterfall" processes. Scrum processes enable organizations to adjust smoothly to rapidly-changing requirements and produce a product that meets evolving business goals.

C. Algorithms

Algorithm: Naïve Bayes

Naïve Bayes is classification algorithm which is based on 'naïve' assumptions that all the features are independent of each other. Naïve Bayes algorithm is faster and easy to implement but it makes assumptions makes it less powerful as compared to other algorithms. It uses probabilistic machine learning model for doing classification.

i. Advantages
- Easy, fast and Scalable.
- Works well with less training data.
- Works well with highdimensions.

ii. Disadvantages
- Carefully handle Laplace Correction.
- Assumption of independent features.

iii. Applications
- Works well with textual data.

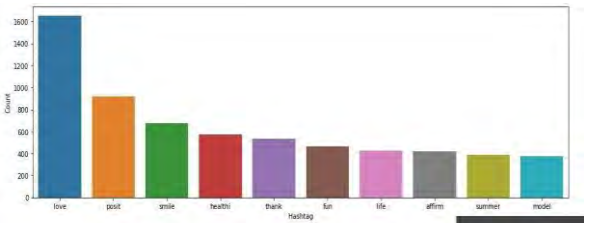


Figure 1: Count of non-racist hashtags

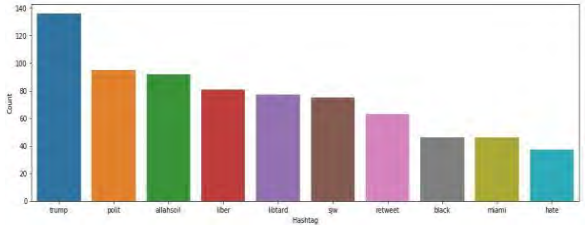


Figure 2: Count of racist hashtags

Algorithm: k-nearest Neighbours

The k-nearest neighbours (KNN) algorithm is one of the simple and easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. KNN works by considering k nearest elements only and takes decision by vote majority. KNN is useful if we do not have and information about distribution of data. KNN is also called as lazy learner as training phase is fast as model is not built, but it computes distance every time when we run the algorithm.

i. Advantages
- Simple and easy to implement.
- No need to build model.
- No parameter tuning required.
- It makes no assumptions.
- Adaptive

ii. Disadvantages
- Slower as data increases.
- Lazy Learner.
- As number of dimensions increases, speed decreases.
- Problem when data is skewed.

iii. Applications
- Previously used in Recommendation Systems.
- Used when no information about distribution of data.
- When we want results faster.

```
#import GridSearchCV
from sklearn.model_selection import GridSearchCV
#In case of classifier like knn the parameter to be tuned is n_neighbors
param_grid = {'n_neighbors':np.arange(1,50)}
knn = KNeighborsClassifier()
knn_cv= GridSearchCV(knn,param_grid,cv=5)
knn_cv.fit(X,y)

print("Best Score:" + str(knn_cv.best_score_))
print("Best Parameters: " + str(knn_cv.best_params_))
```

```
Best Score:0.7721354166666666
Best Parameters: {'n_neighbors': 25}
```

Figure 3: Grid Search

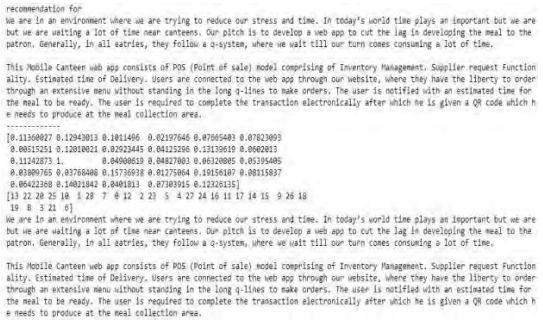


Figure 4: Use of KNN algorithm in Recommendation system

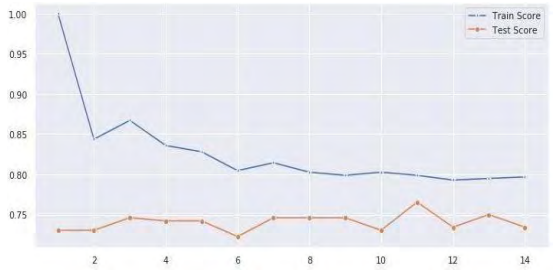


Figure 5: Training vs Testing graph for values of 'k'

Algorithm: Decision Tree Classifier

Decision tree is one of the most widely used classification algorithm. The goal of this algorithm is to create a model that predicts the value of a dependent variable by learning simple decision rules inferred from the data features. It uses a tree like model which represents decision rules.

i. Advantages:

- Easy to visualize decision rule.
- Can work well with noisy data
- Can deal with both numeric and categorial data.
- Almost no hyper-parameter tuning required

ii. Disadvantage:

- Can overfit data easily if not pruned properly.
- Works not so well with skewed data.
- Unstable.
- Biased trees are constructed if some classes dominate.

Algorithm: Logistic regression

Logistic Regression is one of the most popular classification techniques used in Machine Learning. It is primarily used to predict the categorical outcomes may it be binomial or multinomial values. Logistic Regression is termed as a special case of Linear Regression which uses the concept of log odds to determine the relationship of the independent variables with the dependent variables using log it function. Thus, the outcome obtained is actually a probability.

i. Advantages:

- It is computationally efficient.
- Performs better when unrelated and highly correlated features are removed.
- Can be easily combined with other algorithms to form a complex model.

ii. Disadvantages:

- Doesn't perform well with independent variables that aren't correlated with target variable and are much correlated to each other.
- It can't solve nonlinear problems as its decision surface is linear.

iii. Applications:

- It requires the independent variables to be very less dependent on each other ie. the independent variables should not be highly correlated.
- The independent variables are linearly related to the log odds (ie. by logit/sigmoid function).
- The sample size of the observations must be large.

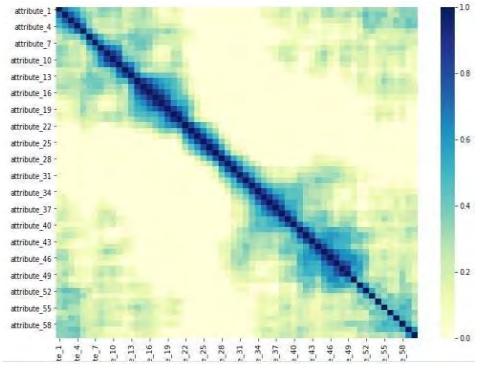


Figure 6: Heatmap of correlation of features

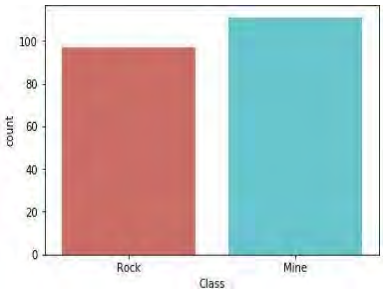


Figure 7: Count of class variables to check if dataset is balanced

Algorithm: Support Vector Machine

SVM is one of the most powerful classification techniques used in Machine Learning. SVM works by finding optimal hyperplane from given dataset which can classify data-points very well. It is primarily used for solving classification problems, but it even works well with regression problems. SVM can deal with non-linearity, small dataset and high dimensions very well.

i. Advantages:

- Can deal with outliers very well if we choose proper regularization parameters (C and Gamma).
- Effective when dimensions are higher.
- Works well even when training examples are less than number of features.
- Can classify linear as well as non linear data

by kernel trick.

- Efficient as we only need support vectors for making prediction.
- Almost works well with all use cases.

ii. Disadvantages

- Chances of overfitting if regularization parameters not chosen properly.
- Training time is more.

iii. Applications

Works almost well with all the use cases, but prominent of the use cases are:

- Image Classification
- Text Classification

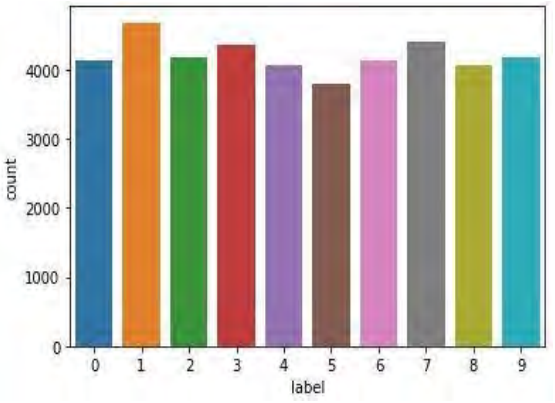


Figure 8: Count of digits to check if data is balanced

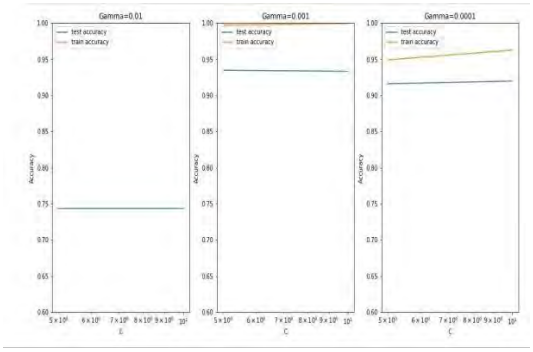


Figure 9: Training vs Testing accuracy at different values of ϒ

D. Observation

Table 1: Observation Table depicting performance measures

Classifier	Dataset	Performance Measures	
Naïve Bayes	Twitter dataset	Precision	82.5%
SVM	Handwriting Recognition dataset	Accuracy	94.4%
Decision Tree	Nursery dataset	Accuracy	86%
KNN	Diabetes dataset	Accuracy	76.5%
		F1 score	83%
		Precision	80%
		Recall	85%
		AUC	0.772
Logistic Regression	Sonar dataset	Accuracy	78%
		F1 score	76.2%
		AUC	0.777

References

Paper in a Proceedings/Transactions/Journal

[1] oyang Dun, Eric Wang, Sagnik Majumder," Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? ", Journal of Machine Learning Research 15 (2014) 3133-3181,2014.

[2] avid J. Dittman, Taghi M. Khoshgoftaar, Amri Napolitano," Selecting the Appropriate Ensemble Learning Approach for Balanced Bioinformatics Data",AAAI Publication,The Twenty-Eighth Internation Flairs Conference, 2018.

[3] asin Uzun," Keyword Extraction Using Naive Bayes", Bilkent University, Department of Computer Science, Turkey,2012

[4] . Domingos and M. Pazzani,"On the optimality of the simple bayesian classifier under zero-one loss", Machine Learning, 29(2/3):103-130, 1997.

[5] arleenKaur and VinitaKumari," Predictive modelling and analytics for diabetes using a machine learning approach", IEEE EMBS Int. Conf. on Biomed. & Health Informat. (BHI), Orlando, FL,2017

Websites

[1] https://towardsdatascience.com/all-about-naive- bayes-8e13cef044cf

[2]https://www.analyticsvidhya.com/blog/2017/08/skilltest-logistic-regression/

[3]https://medium.com/@mohtedibf/in-depth- parameter-tuning-for-knn-4c0

[4] https://towardsdatascience.com/ensemble-methods- bagging-boosting-and-stackin

Content Aggregator Using Web Scraper

Dr. Megharani Patil, Rahul Choudhari, Nitin Dwivedi, Abhishek Gupta
Department of Computer Engineering,
Thakur College Of Engineering and Technology
University of Mumbai
megharani.patil@thakureducation.org,rahulac1599@gmail.com, nitindwivedi88@gmail.com, abhig280031@gmail.com

Abstract - **Content Aggregator is an information hub where a user can access information as per his interests. Content Aggregator targets renowned websites and extract the latest information. Users can access the latest information easily from Content Aggregator.**
There is the huge and rapid growth in the unstructured data every moment. The production and generation of data is predicted to be 44 times in 2020 as compared to the data in 2009. All these figures and statistical data are amazing and growing in exponential pattern. Such data is unstructured in nature which Means the data of different and heterogeneous formats. Web scraping is a term used to describe the use of a program or algorithm to extract and process large amounts of data from the web. The ability to scrape data from the web is a useful skill to have. Let's say you find data from the web, and there is no direct way to download it, web scraping using Python is a skill you can use to extract the data into a useful form that can be imported. The Beautiful Soup package is used to extract data from html files. Recommendation system compares the data collected from users and create a list of items recommended to the user. They are an alternative search algorithm because it helps users to find items and information that they would not find themselves. At present the problem of recommendation systems keeps to himself a lot of interest.

Keywords: Web Scraping, Hybrid Recommendation, Beautiful Soup

I. INTRODUCTION:

A content aggregator is an individual or organization that gathers Web content from different online sources for reuse or resale. There are two kinds of content aggregators: (1) those who simply gather material from various sources for their Web sites, and (2) those who gather and distribute content to suit their customer's needs.

.

II. RELATED WORK: -

A Dive into Web Scraper World (IEEE) HMR Institute of Technology Delhi by Deepak Kumar Mahto HMR Institute of Technology and Management, Delhi. We studied about Data Extraction Using Web Crawler, some legal issues and Designing Principles. This involves various methods and techniques in Web Scraping. Gaps Identified in our project by this paper are unable to resolve absolute URL and handling of redirect loops. [1] Information Extraction Using Web Usage Mining and Web Scraping (IEEE) University of Information Technology New Delhi by Sanjay Kumar Malik SAM Rizvi Extraction of useful data from discovery patterns using web mining. Conversion of unstructured data into a designated format. Gaps identified are Web scraping is limiting as you must know the html formatting, and we need to find a way around that.

[2]. A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System by JonAtle Gulla and R. Cenk Erdur. News recommendation have specific challenges when compared to the other domains. From a technical point of view there are many different methods to build a recommender system. Thus, while general methods are used in news recommendation, researchers also need some new methods to make proper news recommendations. In this paper we present the different approaches to news recommender systems and the challenges of news recommendation. Gaps identified are Key challenge is predicting the future interests of users for better recommendations News items should be structured and machine readable.

[3] We all use Web Browser to extract the needed information from the Web Site, if you think this is the only way to access information from the internet then you are missing out on a huge range of available possibilities. Web Scraper is one of those possibilities, in which we are accessing the information on the internet using some programs and pre written libraries. Most websites do not offer the functionality to save a copy of the data which they display to your computer. The only option then is to manually copy and paste the data displayed by the website in your browser to a local file in your computer - a very tedious job which can take many hours or sometimes days to complete.

Over the last few years consumption of news articles has shifted more and more from the written versions towards the web. Internet devices like Desktop, Laptop or even a Mobile Phone became more powerful, with larger screens and connected to the Internet, have had a great influence on this paradigm change. A critical problem associated with online news is related to the fact that the large number of daily articles can be overwhelming to the users. Recommendation services can largely improve the efficiency and accuracy of acquired information. These systems are designed to filter critical news, key events and

Formatted: All caps

Formatted: List Bullet, Centered, Space Before: 0 pt, Numbered + Level: 1 + Numbering Style: I, II, III, … + Start at: 1 + Alignment: Right + Aligned at: 0.63 cm + Indent at: 1.27 cm, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border), Bar : (No border)

meaningful items that might be of interest to a reader. Recommendation system model keeps the count of user interests and accordingly shows relevant information

~~II.~~III. TECHNOLOGY USED: -

Flask: - Flask (source code) is a Python web framework built with a small core and easy-to-extend philosophy. Flask is considered more Python than the Jingo web framework because in common situations the equivalent Flask web application is more explicit. Flask is also easy to get started with as a beginner because there is little boilerplate code for getting a simple app up and running.

Web Scraping: - Web scraping (web harvesting or web data extraction) is a computer software technique of extracting information from websites.HTML parsing is easy in Python, especially with the help of the Beautiful Soup library.

Beautiful Soup: Beautiful Soup is a Python package for parsing HTML and XML documents. It creates parse trees that is helpful to extract the data easily. There are mainly two ways to extract data from a website:

1. Use the API of the website (if it exists). For example, Facebook has the Facebook Graph API which allows retrieval of data posted on Facebook.
2. Access the HTML of the webpage and extract useful information/data from it. This technique is called web scraping or web harvesting or web data extraction.

~~III.~~IV. IMPLEMENTATION DETAIL

Step 1 Scraping raw data:

Our target websites include data across various fields such as sports, news, etc. To scrape such sites, we take the help of Beautiful Soup Library and Requests Package present in Python Library. A target site is requested and a beautiful soup object is created with the help of data request. Various beautiful soup methods are applied to filter data and acquire desired result. Output can be stored in various formats such as .csv, json format, directly into

A database. The stored data is displayed on aggregator site as per user's recommendations and content viewing. Step 2 Model selection:

Our proposed system for recommendation system will work on hybrid model which is the combination of collaborative filtering and content-based filtering. These model will be Machine learning based and use linear regression as learning algorithm. Consider an example where 5 users has given rating to some of the articles, each article has a particular set of attributes which shows what kind of article it is. Let's assume that attribute 1 represent sports news and attribute 2 represents Entertainment news taking example of article 1 which has attribute 1 value of

0.99 And attribute 2 value of 0.02 which mean article 1 is mostly a sports based article.

	User A	User B	User C	User D	User E	Attribute 1	Attribute 2
Article 1	5	5	5	0	0	0.99	0.02
Article 2	5	?	5	0	?	1.50	0.01
Article 3	?	5	4	?	1	1.00	0.1
Article 4	0	0	?	5	5	0.09	1.3
Article 5	1	0	0	5	4	0.2	1.05

Note: Here the '?' indicates that the user did not encounter the article yet.

Fig 3.1 Example of data model

Content based approach

Content based recommendation identifies the similar items to those that the target user has liked in the past.

In the proposed model the content-based approach will try to predict the rating of the article which is not encounter yet with the help of attributes of the article.

Collaborative filtering approach

Collaborative filtering identifies the similarity between two or group of users which have similar interests. In collaborative filtering the model will predict the attributes value of articles with the help of ratting of different usersAnd recommend those articles which are very similar to user interest

Hybrid model

The hybrid model is derived from the above mentioned two methods. Each methods have its own advantages and disadvantages, to make recommendation system more reliable we use hybrid model. Hybrid model will try to work like both approaches at the same time i.e. it will try to compute both the new ratting of article for the different users and the attribute of new article. The Hybrid model will be trained using linear regression, like other linear regression model our hybrid model will basically compute the error cost and try to minimize the cost, the cost function is also derived for both models to make hybrid model more reliable

Step 3 Algorithm:

r(i,j) = 1 if user j has rated article i (0 otherwise)

y(i,j) = rating by user j on article I ɵ(j) = parameter vector for user j x(i) = feature vector for article i Cost function:

Minimize x(1), . . . , x(n) and ɵ(1), . . . , ɵ(m) Simultaneously:

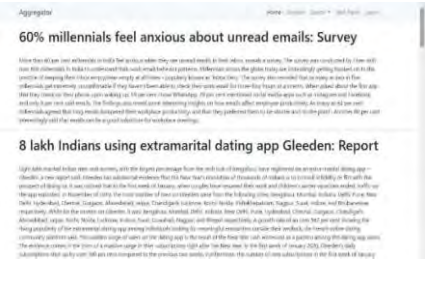


Fig 3.2 Cost function

Our goal is to minimize cost function these can be achieved by gradient descent by updating and x at every iteration.

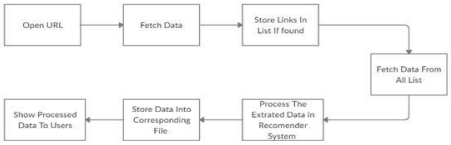


Fig 3.3 Weight update using gradient descent

Sample Code:-
Scraping Function on Target Site eg: FIFA World Rankings

From bs4 import Beautiful Soup import csv
Import requests def scrape ():
url="https://www.fifa.com/fifa-world- ranking/rankin-table/men/EON/wepAnan/ranking/men"f = csv.writer(opencsv/fifa rankings-men.csv', 'W'))
Html = BeautifulSoup(data.text, 'html.parser') x =html.find all(class='fi-t_ntext') f.writerow(['Rank', 'Name']) extracted_records = []

For i in x:
title =
i.text
record
= {
'title': title,
'rank': c
f.writerow([c,
title]])| c += 1
extracted_records.append (record)
Return (extracted records)

~~IV.~~ V. BLOCK DIAGRAM/FLOWCHART:

5.1 Block Diagram

The below fig 5.1 shows the principal component and function in news content aggregator and shows the flow of the process

$$x_k^{(i)} = x_k^{(i)} - \alpha\left(\sum_{j:r(i,j)=1}((\theta^{(j)})^T x^{(i)} - y^{(i,j)})\theta_k^{(j)} + \lambda x_k^{(i)}\right)$$

$$\theta_k^{(j)} = \theta_k^{(j)} - \alpha\left(\sum_{i:r(i,j)=1}((\theta^{(j)})^T x^{(i)} - y^{(i,j)})\theta_k^{(j)} + \lambda\theta_k^{(j)}\right)$$

Fig 5.1 Block Diagram

5.2 Web scraping Activity Diagram

The Activity diagrams fig 3.6.6 shows the graphical representations of workflows of
Stepwise activities and actions with support for choice, iteration and concurrency of content aggregator.

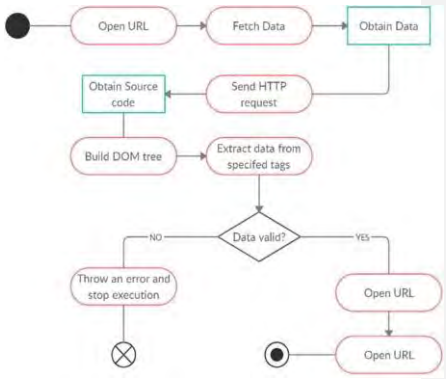


Fig 5.2 Web Scraping Activity Diagram6.
Result:

~~V.~~ VI. EXPECTED OUTCOME:

The expected Outcome of the project are:
1. Content Aggregator website which is an Information hub where a user can access Information as per his interests.
2. Users can access the latest information Easily from Content Aggregator.
3. The latest and trending news is Recommended to the user as per his Subscriptions and feed history.
Fig.

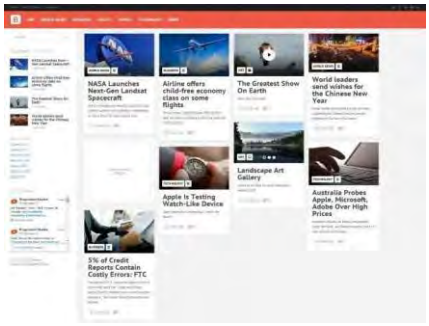


Fig. 6.1 Expected User Interface

- Snapshots:

From the data extracted from the target URL
The contents are stored in a CSV file and the
Data from the CSV file is displayed on the
Website

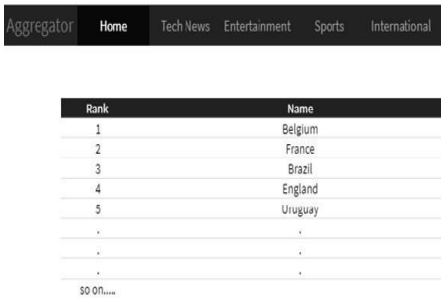


Fig. 6.2.FIFA Rankings

~~VI.~~ VII. CONCLUSION

Content Aggregator Using Web Scraper thus
Provides insight of extracting the information from
Websites in cases where it is difficult to fetch the
Information due to structure or format of the web page
And provide crisp look of those data on a single web
Page in more elaborative and effective way. Thus user
Accessing the web page will get only relevant
Information of every renowned web pages thus save
User time of searching all of the pages. The interest on
Online newspapers has been growing significantly over
The past years. In order to present the most relevant
News articles to users, different recommendation
Systems have been made available using various
Techniques in order to make access to large amounts of
Information more efficient.
Mobility, social networking and the large
Number of news providers, bring new challenges but
Also new opportunities to enhance access to information
And improve the user experience when browsing
Newspapers. Our hybrid recommendation system for
News in a mobile environment that takes into
Consideration aspects that reflect this new paradigm and
Provides user's customized feed and as per his activity
Using a hybrid recommender system of content and
Collaborative method.

References

[1]"A Dive into Web Scraper World (IEEE)" HMR Institute of Technology Delhi 2016 International Conference on Computing for Sustainable Global Development
[2] "A Novel Approach to Web Scraping Technology "International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 5, MAY 2015 ISSN: 2277 128 X[3] "A hybrid recommendation system for news in a
mobile environment" Paula Viana and Márcio Soares, Conference Paper · June 2016 DOI: 10.1145/2912845.2912852
[4] "An Improved Approach for Movie Recommendation System" Shreya Agrawal and Pooja Jain CSE SVITS Indore, MP, India, International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017)
[5] "Information Extraction Using Web Usage Mining and Web Scraping.(IEEE)" University of Information Technology New Delhi 2011 International Conference on Computational Intelligence and Communication Systems
[6] "Conceptual Approach for the Development of Web Scraping Application for Tracking Information" University of National and World Economy 2017 Economic Alternatives, Issue #3 -Plamen Milev
[7] "Methods and Tools for Building Recommender Systems" Yuri Stekh, Mykhoylo Lobur, Vitalij Artsibasov, Vitalij Chystyak, CADSM 2015, 24-27 February, 2015, Polyana-Svalyava (Zakarpattya), UKRAINE
[8] "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System" Geetha G, Safa M, Fancy C, Saranya D,National Conference on Mathematical Techniques and its Applications (NCMTA 18)
[9] "A Survey on Challenges and Methods in News Recommendation" Ozlem Ozgobek, Jon Atle Gulla and R. Cenk Erdur.
[10] "Web Information Retrieval Using Python and BeautifulSoup" International Journal for Research in Applied Science & Engineering Technology (IJRASET)Volume 4 Issue VI, June 2016, IC Value: 13.98, ISSN: 2321-9653.
[11] Ryan Mitchell. "Web Scraping with Python: Collecting Data from the Modern Web", O'Reilly Media, Inc. publishers, 1st Edition, 2015-06-10.
[12] Katharine Jarmul and Richard Lawson. "Python Web Scraping", Packt Publishing, 2nd Edition, May 2017.
[13] Python Web Scraping, available at https://realpython.com/tutorials/web-scraping accessed on Sept 2019.
[14] Web Scraping using BeautifulSoup,available at https://www.datacamp.com/community/tutorials/webscraping-using-python accessed on Oct 2019.

Content Aggregator Using Web Scraper

Dr. Megharani Patil, Rahul Choudhari, Nitin Dwivedi, Abhishek Gupta
Department of Computer Engineering,
Thakur College Of Engineering and Technology
University of Mumbai
megharani.patil@thakureducation.org,rahulac1599@gmail.com, nitindwivedi88@gmail.com, abhig280031@gmail.com

Abstract - **Content Aggregator is an information hub where a user can access information as per his interests. Content Aggregator targets renowned websites and extract the latest information. Users can access the latest information easily from Content Aggregator.**
There is the huge and rapid growth in the unstructured data every moment. The production and generation of data is predicted to be 44 times in 2020 as compared to the data in 2009. All these figures and statistical data are amazing and growing in exponential pattern. Such data is unstructured in nature which Means the data of different and heterogeneous formats. Web scraping is a term used to describe the use of a program or algorithm to extract and process large amounts of data from the web. The ability to scrape data from the web is a useful skill to have. Let's say you find data from the web, and there is no direct way to download it, web scraping using Python is a skill you can use to extract the data into a useful form that can be imported. The Beautiful Soup package is used to extract data from html files. Recommendation system compares the data collected from users and create a list of items recommended to the user. They are an alternative search algorithm because it helps users to find items and information that they would not find themselves. At present the problem of recommendation systems keeps to himself a lot of interest.

Keywords: Web Scraping, Hybrid Recommendation, Beautiful Soup

I. INTRODUCTION:

A content aggregator is an individual or organization that gathers Web content from different online sources for reuse or resale. There are two kinds of content aggregators: (1) those who simply gather material from various sources for their Web sites, and (2) those who gather and distribute content to suit their customer's needs.

.

II. RELATED WORK: -

A Dive into Web Scraper World (IEEE) HMR Institute of Technology Delhi by Deepak Kumar Mahto HMR Institute of Technology and Management, Delhi. We studied about Data Extraction Using Web Crawler, some legal issues and Designing Principles. This involves various methods and techniques in Web Scraping. Gaps Identified in our project by this paper are unable to resolve absolute URL and handling of redirect loops. [1] Information Extraction Using Web Usage Mining and Web Scraping (IEEE) University of Information Technology New Delhi by Sanjay Kumar Malik SAM Rizvi Extraction of useful data from discovery patterns using web mining. Conversion of unstructured data into a designated format. Gaps identified are Web scraping is limiting as you must know the html formatting, and we need to find a way around that.

[2]. A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System by JonAtle Gulla and R. Cenk Erdur. News recommendation have specific challenges when compared to the other domains. From a technical point of view there are many different methods to build a recommender system. Thus, while general methods are used in news recommendation, researchers also need some new methods to make proper news recommendations. In this paper we present the different approaches to news recommender systems and the challenges of news recommendation. Gaps identified are Key challenge is predicting the future interests of users for better recommendations News items should be structured and machine readable.

[3] We all use Web Browser to extract the needed information from the Web Site, if you think this is the only way to access information from the internet then you are missing out on a huge range of available possibilities. Web Scraper is one of those possibilities, in which we are accessing the information on the internet using some programs and pre written libraries. Most websites do not offer the functionality to save a copy of the data which they display to your computer. The only option then is to manually copy and paste the data displayed by the website in your browser to a local file in your computer - a very tedious job which can take many hours or sometimes days to complete.

Over the last few years consumption of news articles has shifted more and more from the written versions towards the web. Internet devices like Desktop, Laptop or even a Mobile Phone became more powerful, with larger screens and connected to the Internet, have had a great influence on this paradigm change. A critical problem associated with online news is related to the fact that the large number of daily articles can be overwhelming to the users. Recommendation services can largely improve the efficiency and accuracy of acquired information. These systems are designed to filter critical news, key events and

Formatted: All caps

Formatted: List Bullet, Centered, Space Before: 0 pt, Numbered + Level: 1 + Numbering Style: I, II, III, … + Start at: 1 + Alignment: Right + Aligned at: 0.63 cm + Indent at: 1.27 cm, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border), Bar : (No border)

meaningful items that might be of interest to a reader. Recommendation system model keeps the count of user interests and accordingly shows relevant information

~~II.~~III. TECHNOLOGY USED: -

Flask: - Flask (source code) is a Python web framework built with a small core and easy-to-extend philosophy. Flask is considered more Python than the Jingo web framework because in common situations the equivalent Flask web application is more explicit. Flask is also easy to get started with as a beginner because there is little boilerplate code for getting a simple app up and running.

Web Scraping: - Web scraping (web harvesting or web data extraction) is a computer software technique of extracting information from websites.HTML parsing is easy in Python, especially with the help of the Beautiful Soup library.

Beautiful Soup: Beautiful Soup is a Python package for parsing HTML and XML documents. It creates parse trees that is helpful to extract the data easily. There are mainly two ways to extract data from a website:

1. Use the API of the website (if it exists). For example, Facebook has the Facebook Graph API which allows retrieval of data posted on Facebook.
2. Access the HTML of the webpage and extract useful information/data from it. This technique is called web scraping or web harvesting or web data extraction.

~~III.~~IV. IMPLEMENTATION DETAIL

Step 1 Scraping raw data:

Our target websites include data across various fields such as sports, news, etc. To scrape such sites, we take the help of Beautiful Soup Library and Requests Package present in Python Library. A target site is requested and a beautiful soup object is created with the help of data request. Various beautiful soup methods are applied to filter data and acquire desired result. Output can be stored in various formats such as .csv, json format, directly into

A database. The stored data is displayed on aggregator site as per user's recommendations and content viewing. Step 2 Model selection:

Our proposed system for recommendation system will work on hybrid model which is the combination of collaborative filtering and content-based filtering. These model will be Machine learning based and use linear regression as learning algorithm. Consider an example where 5 users has given rating to some of the articles, each article has a particular set of attributes which shows what kind of article it is. Let's assume that attribute 1 represent sports news and attribute 2 represents Entertainment news taking example of article 1 which has attribute 1 value of

0.99 And attribute 2 value of 0.02 which mean article 1 is mostly a sports based article.

	User A	User B	User C	User D	User E	Attribute 1	Attribute 2
Article 1	5	5	5	0	0	0.99	0.02
Article 2	5	?	5	0	?	1.50	0.01
Article 3	?	5	4	?	1	1.00	0.1
Article 4	0	0	?	5	5	0.09	1.3
Article 5	1	0	0	5	4	0.2	1.05

Note: Here the '?' indicates that the user did not encounter the article yet.

Fig 3.1 Example of data model

Content based approach

Content based recommendation identifies the similar items to those that the target user has liked in the past.

In the proposed model the content-based approach will try to predict the rating of the article which is not encounter yet with the help of attributes of the article.

Collaborative filtering approach

Collaborative filtering identifies the similarity between two or group of users which have similar interests. In collaborative filtering the model will predict the attributes value of articles with the help of ratting of different usersAnd recommend those articles which are very similar to user interest

Hybrid model

The hybrid model is derived from the above mentioned two methods. Each methods have its own advantages and disadvantages, to make recommendation system more reliable we use hybrid model. Hybrid model will try to work like both approaches at the same time i.e. it will try to compute both the new ratting of article for the different users and the attribute of new article. The Hybrid model will be trained using linear regression, like other linear regression model our hybrid model will basically compute the error cost and try to minimize the cost, the cost function is also derived for both models to make hybrid model more reliable

Step 3 Algorithm:

r(i,j) = 1 if user j has rated article i (0 otherwise)

y(i,j) = rating by user j on article I ɵ(j) = parameter vector for user j x(i) = feature vector for article i Cost function:

Minimize x(1), . . . , x(n) and ɵ(1), . . . , ɵ(m) Simultaneously:

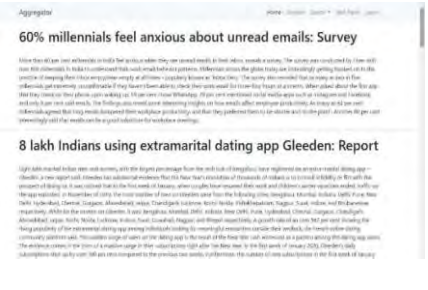


Fig 3.2 Cost function

Our goal is to minimize cost function these can be achieved by gradient descent by updating and x at every iteration.

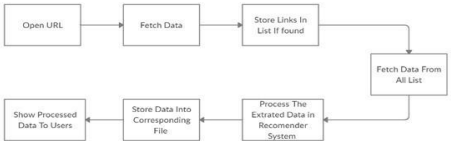


Fig 3.3 Weight update using gradient descent

Sample Code:-
Scraping Function on Target Site eg: FIFA World Rankings

From bs4 import Beautiful Soup import csv
Import requests def scrape ():
url="https://www.fifa.com/fifa-world- ranking/rankin-table/men/EON/wepAnan/ranking/men"f = csv.writer(opencsv/fifa rankings-men.csv', 'W'))
Html = BeautifulSoup(data.text, 'html.parser') x =html.find all(class='fi-t_ntext') f.writerow(['Rank', 'Name']) extracted_records = []

For i in x:
title =
i.text
record
= {
'title': title,
'rank': c
f.writerow([c,
title]])| c += 1
extracted_records.append (record)
Return (extracted records)

~~IV.~~ V. BLOCK DIAGRAM/FLOWCHART:

5.1 Block Diagram

The below fig 5.1 shows the principal component and function in news content aggregator and shows the flow of the process

$$x_k^{(i)} = x_k^{(i)} - \alpha\left(\sum_{j:r(i,j)=1}((\theta^{(j)})^T x^{(i)} - y^{(i,j)})\theta_k^{(j)} + \lambda x_k^{(i)}\right)$$

$$\theta_k^{(j)} = \theta_k^{(j)} - \alpha\left(\sum_{i:r(i,j)=1}((\theta^{(j)})^T x^{(i)} - y^{(i,j)})\theta_k^{(j)} + \lambda\theta_k^{(j)}\right)$$

Fig 5.1 Block Diagram

5.2 Web scraping Activity Diagram

The Activity diagrams fig 3.6.6 shows the graphical representations of workflows of
Stepwise activities and actions with support for choice, iteration and concurrency of content aggregator.

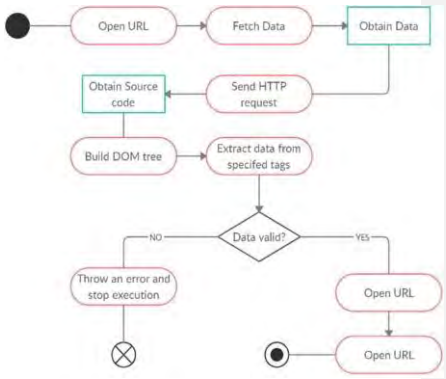


Fig 5.2 Web Scraping Activity Diagram6.
Result:

~~V.~~ VI. EXPECTED OUTCOME:

The expected Outcome of the project are:
1. Content Aggregator website which is an Information hub where a user can access Information as per his interests.
2. Users can access the latest information Easily from Content Aggregator.
3. The latest and trending news is Recommended to the user as per his Subscriptions and feed history.
Fig.

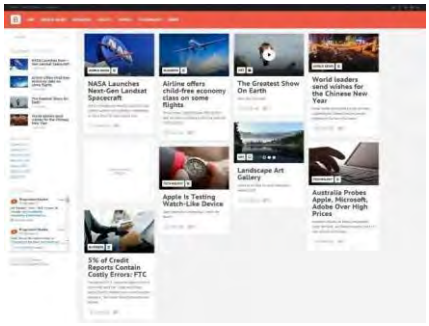


Fig. 6.1 Expected User Interface

- Snapshots:

From the data extracted from the target URL
The contents are stored in a CSV file and the
Data from the CSV file is displayed on the
Website

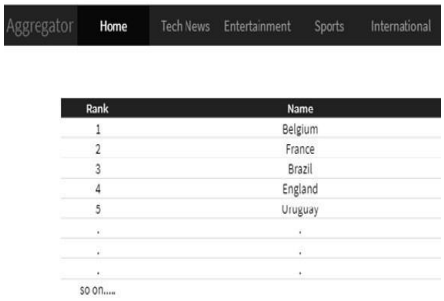


Fig. 6.2.FIFA Rankings

~~VI.~~ VII. CONCLUSION

Content Aggregator Using Web Scraper thus
Provides insight of extracting the information from
Websites in cases where it is difficult to fetch the
Information due to structure or format of the web page
And provide crisp look of those data on a single web
Page in more elaborative and effective way. Thus user
Accessing the web page will get only relevant
Information of every renowned web pages thus save
User time of searching all of the pages. The interest on
Online newspapers has been growing significantly over
The past years. In order to present the most relevant
News articles to users, different recommendation
Systems have been made available using various
Techniques in order to make access to large amounts of
Information more efficient.
Mobility, social networking and the large
Number of news providers, bring new challenges but
Also new opportunities to enhance access to information
And improve the user experience when browsing
Newspapers. Our hybrid recommendation system for
News in a mobile environment that takes into
Consideration aspects that reflect this new paradigm and
Provides user's customized feed and as per his activity
Using a hybrid recommender system of content and
Collaborative method.

References

[1]"A Dive into Web Scraper World (IEEE)" HMR Institute of Technology Delhi 2016 International Conference on Computing for Sustainable Global Development
[2] "A Novel Approach to Web Scraping Technology "International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 5, MAY 2015 ISSN: 2277 128 X[3] "A hybrid recommendation system for news in a
mobile environment" Paula Viana and Márcio Soares, Conference Paper · June 2016 DOI: 10.1145/2912845.2912852
[4] "An Improved Approach for Movie Recommendation System" Shreya Agrawal and Pooja Jain CSE SVITS Indore, MP, India, International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017)
[5] "Information Extraction Using Web Usage Mining and Web Scraping.(IEEE)" University of Information Technology New Delhi 2011 International Conference on Computational Intelligence and Communication Systems
[6] "Conceptual Approach for the Development of Web Scraping Application for Tracking Information" University of National and World Economy 2017 Economic Alternatives, Issue #3 -Plamen Milev
[7] "Methods and Tools for Building Recommender Systems" Yuri Stekh, Mykhoylo Lobur, Vitalij Artsibasov, Vitalij Chystyak, CADSM 2015, 24-27 February, 2015, Polyana-Svalyava (Zakarpattya), UKRAINE
[8] "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System" Geetha G, Safa M, Fancy C, Saranya D,National Conference on Mathematical Techniques and its Applications (NCMTA 18)
[9] "A Survey on Challenges and Methods in News Recommendation" Ozlem Ozgobek, Jon Atle Gulla and R. Cenk Erdur.
[10] "Web Information Retrieval Using Python and BeautifulSoup" International Journal for Research in Applied Science & Engineering Technology (IJRASET)Volume 4 Issue VI, June 2016, IC Value: 13.98, ISSN: 2321-9653.
[11] Ryan Mitchell. "Web Scraping with Python: Collecting Data from the Modern Web", O'Reilly Media, Inc. publishers, 1st Edition, 2015-06-10.
[12] Katharine Jarmul and Richard Lawson. "Python Web Scraping", Packt Publishing, 2nd Edition, May 2017.
[13] Python Web Scraping, available at https://realpython.com/tutorials/web-scraping accessed on Sept 2019.
[14] Web Scraping using BeautifulSoup,available at https://www.datacamp.com/community/tutorials/webscraping-using-python accessed on Oct 2019.

Malaria Detection Using Deep Learning

Manjyotsingh Nanra, Niharika Chaturvedi, Vishal Nadar, Veena Kulkarni
Computer Engineering TCET
Mumbai
nanramanjyot@gmail.com, niharikachaturvedi42@gmail.com, vishalnadar98@gmail.com, Indiaveena.kulkarni@thakureducation.org

Abstract— **Malaria is mosquito-borne blood disease caused by parasites of the genus Plasmodium. According to the World Malaria Report 2016, an estimated 3.2 billion people in 95 countries and territories are at risk of being infected with malaria and developing disease, and 1.2 billion are at high risk (>1 in 1000 chance of getting malaria in a year). There were about 214 million cases of malaria globally in 2016 and about 438,000 malaria deaths. we construct a new image processing system for detection and quantification of plasmodium parasites in blood smear slide. Conventional diagnostic tool for malaria is the examination of stained blood cell of patient in microscope. The blood to be tested is placed in a slide and is observed under a microscope to count the number of infected RBC. The blood to be tested is placed in a slide and is observed under a microscope to count the number of infected RBC. Based on the guidelines from the WHO protocol, this procedure involves intensive examination of the blood smear at a 100X magnification, where people manually count red blood cells that contain parasites out of 5000 cells. An expert technician is involved in the examination of the slide with intense visual and mental concentration. This is tiresome and time- consuming process. Deep Learning models, or to be more specific, Convolutional Neural Networks (CNNs) have proven to be really effective in a wide variety of computer vision tasks. The proposed work is to develop automated system for detection of Malaria Parasites (MP) in cell images. This system is primarily based on Image processing of Cell Images acquired through digital microscope and classify it using Convolution Neural Network (CNN). This is to create an automatic system which will be able to detect and classify infected RBC cells. The classification of RBC will be entirely dependent on image processing and CNN. The image processing will have removal of impurities, noises and WBCs from image acquired through digital microscope and from remaining enhanced image, extract parasites. Then CNN will be used to identify species ofParasites.**

I. INTRODUCTION

Malaria is caused by protozoan parasites of the genus Plasmodium that are transmitted through the bites of infected female Anopheles mosquitoes and that infect the red blood cells. Most deaths occur among children in Africa, where a child dies almost every minute from malaria, and where malaria is a leading cause of childhood neuro- disability. According to the World Malaria Report 2016,anestimated 3.2 billion people in 95 countries and territories are at risk of being infected with malaria and developing disease, and 1.2 billion are at high risk (>1 in 1000 chance of getting malaria in a year). There were about 214 million cases of malaria globally in 2016 and about 438,000 malaria deaths. The burden was heaviest in the African region, where an estimated 92%2 of all malaria deaths occurred, and in children aged under 5 years, who accounted for more than two thirds of all deaths. Typical symptoms of malaria include fever, fatigue, headaches, and, in severe cases, seizures and coma, leading todeath.

II. LITERATUREREVIEW

H.A Mohammed reviewed various techniques for preprocessing. Deepa. A. Kureret al. [4], proposed a new approach for low-level image processing - SUSAN (Smallest Unvalued segment assimilating nucleus) Principle, which performs Edge and Corner detection. Image features based on color, texture and the geometry of the cells and parasites are generated, as well as features that make use of a priori knowledge of the classification problem and mimic features used by human technicians. A two-stage tree classifier distinguishes between true and false positives, and then diagnoses the four types of the infection. The algorithm detects the species of parasite with a sensitivity of 99% and a positive predictive value of 90-92% Daniel MaitethiaMemeu et al. [5] proposed a method for detection of plasmodium parasites from images of thin blood smears. The method is based on Artificial Neural Net-work (ANN) for testing the presence of plasmodium parasites in thin blood smear images. Pre-processing, features extraction are implemented and eventually diagnosis was made based on the features extracted from the images. Classification accuracy of 95.0% in detection of infected erythrocyte was achieved with respect to results obtained by expert microscopes. S. Kareem et al. [6] proposed a novel idea to identify the total number of red blood cells (RBCs) as well as their location in a Giemsa stained thin blood film images. The method utilizes basic knowledge on cell structure and brightness of the components due to Giemsa staining of the sample and detects and locates the RBCs in the image.Fordetection of Malaria, this study presented a simple approach based on the shape and size of each RBC by calculating its roundness using an appropriate threshold. The classification of malaria parasite is implemented using NCC that does not require training like the ANN.

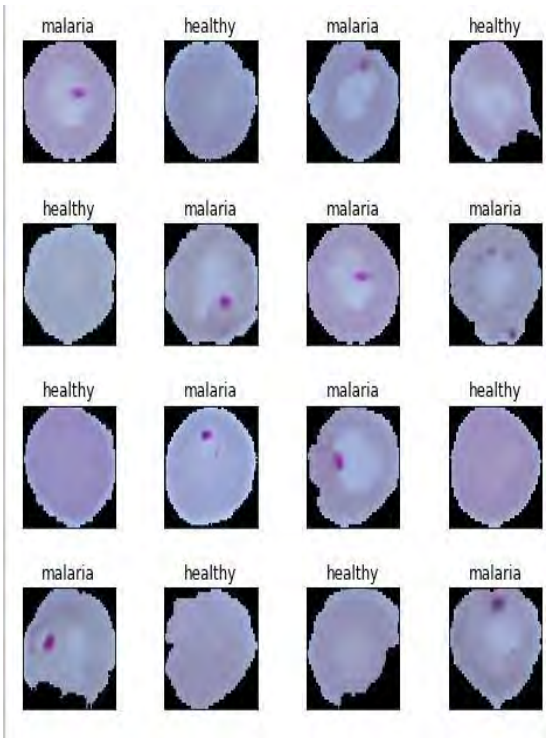
III. MOTIVATION

Conventional diagnostic tool for malaria is the examination of stained blood cell of patient in microscope. The blood to be tested is placed in a slide and is observed under a microscope to count the number of infected RBC. Based on the guidelines from the WHO protocol, this procedure involves intensive examination of the blood smear at a 100X magnification, where people manually count red blood cells that contain parasites out of 5000 cells. An expert technician is involved in the examination of the slide with intense visual and mental concentration. This is tiresome and time- consumingprocess.

IV. METHODOLOGY

A. ImageAcquisition

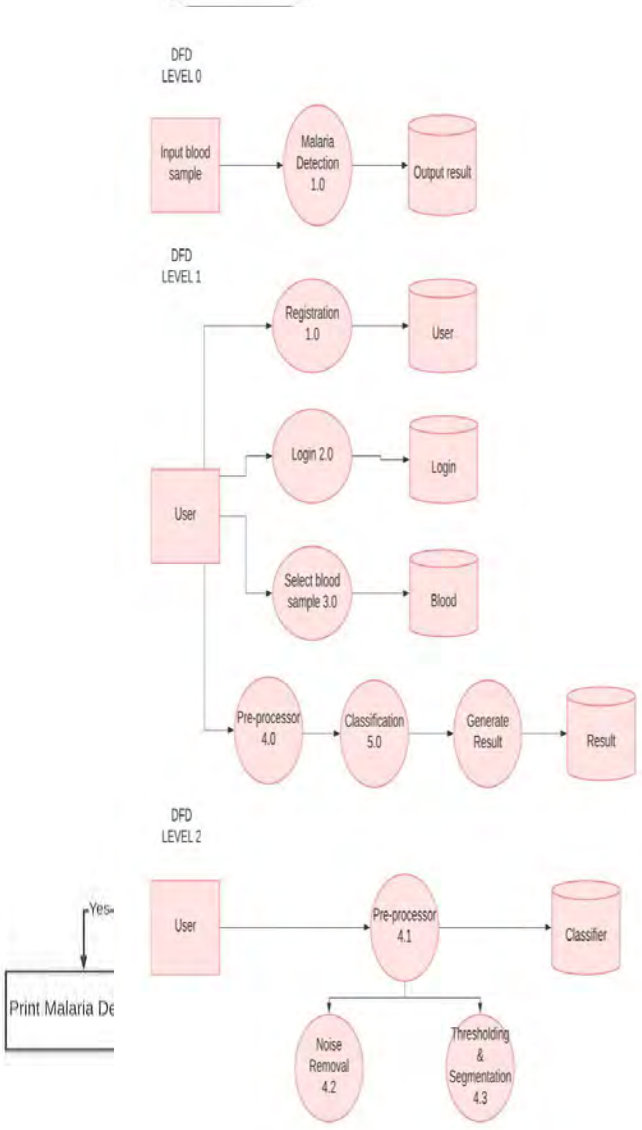
Images of segmented blood cells acquired from Kaggle[11] was used for training of our model. There are about 26000 images pre- classified into respective classes. The classes being Uninfected and Parasitized. Taking help of these datasets we build our model. Images are of followingformat



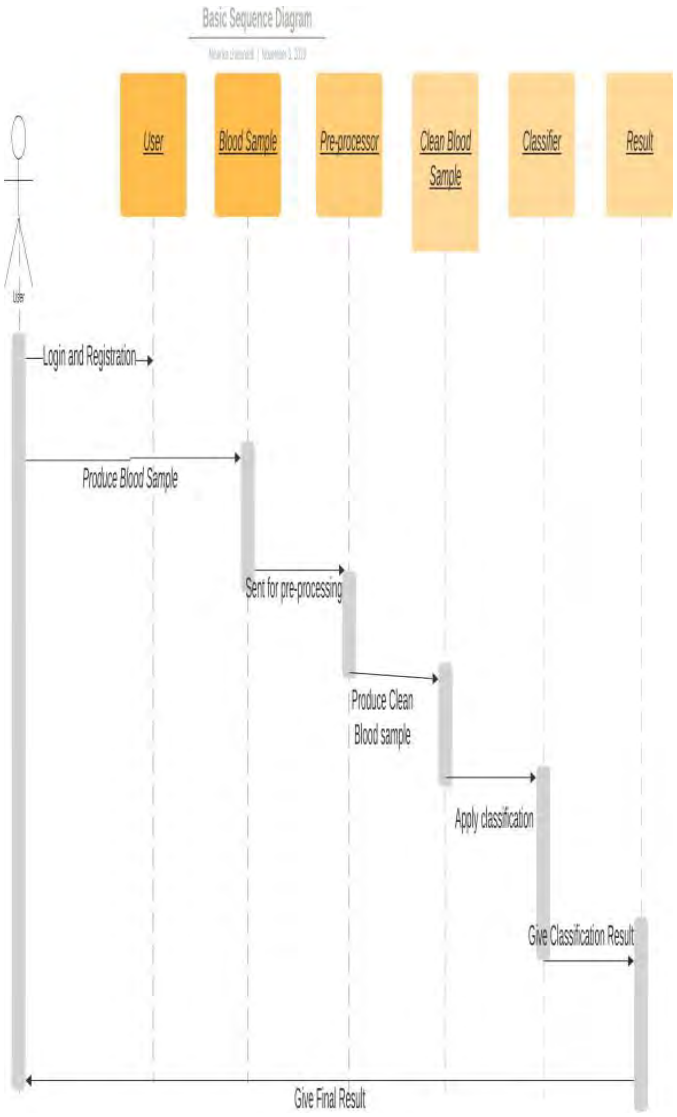
B. ProposedArchitecture

C. Flowchart

D. FunctionalModel



E. SequenceDiagram



V. APPLICATION OFPROJECT

Conventional diagnostic tool for malaria is the examination of stained blood cell of patient in microscope. The blood to be tested is placed in a slide and is observed under a microscope under 100x zoom to count the number of infected RBC. An expert technician is involved in the examination of the slide with intense visual and mental concentration. This is tiresome and time- consuming process. Hence by using image recognition along with machine learning it would become easy to detect malaria accurately and within less time, especially in areas with less resources or with less experiencedpeople.

ACKNOWLEDGMENT

We sincerely thank our guide Mrs. Veena Kulkarni for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work. We thank the HOD, Dr. SheetalRathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support

VI. RESULT AND ANALYSIS

A. Result

We were able to use image processing to get image in desired for. For malaria parasite detection we used Convolutional Neural Network (CNN). We were able to achieve accuracy of 95% with available dataset

B. FurtherWork

Further work needs to be done is to acquire dataset which isn't pre-processed and apply our algorithms on those datasets. Also, accuracy is to be improved so that this model can be used for practical purposes. It should also enable for early detection of malaria and detect various malaria parasite

VII. References

[1] WHO, "fact sheet: world malaria report 2016," in world health organization, world health organization, 2016. [Online]. Available: http://www.Who.Int/malaria/media/world-malaria-report-2016/en. Accessed: jan. 25, 2017

[2] N. A. Khan, H. Pervaz, A. Latif, and A. Musharaff, "*unsupervised identification of malaria parasites using computer vision*,"2017.

[3] ShipraSaraswat, UtkarshAvasthi, NeetuFaujdar, *"Malarial Parasites Detection in RBC Using Image Processing",* 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions*),2017*

[4] Deepa.A.Kurer, Vineeta.P.Gejji, "Detection of Malarial Parasites in Blood Images," International Journal of Engineering Science and Innovative Technology (IJESIT),Volume 3, Issue 3, May2014.

[5] Daniel MaitethiaMemeu et al, "Detection of plasmodium parasites from images of thin blood smears," International Journal of Engineering Science and Innovative Technology (IJESIT), University of Nairobi, Nairobi, Kenya , Open Journal of Clinical Diagnostics, 3, 183-194,2013.d

[6] Saumya Kareem Reni, "Automated Low-Cost Malaria Detection System in Thin Blood Slide Images Using Mobile Phones," Ph.D. dissertation, Dept. Sci and Tech., WestminsterUniv.,uk,2014.

[7] CDC, "Malaria parasites," CDC, 2016. [Online]. Available:

https://www.cdc.gov/malaria/about/biology/parasites.html. Accessed: Jan. 21,2016.

[8] learn zoology. (2014, March 15). Plasmodium Vivax [Online]. Availablehttps://learnzoology.files.wordpress.com/2014/03/p-vivax- schizont.jpg.

[9] K. Prasad, et al"Image Analysis Approach for Development of a Decision Support System for Detection of Malaria Parasites in Thin Blood Smear Images," Journal of Digital Imaging, vol. Volume 25, no. Issue 4, pp. 542–549, Aug.2012.

[10] Hassan Abdelrhman Mohammed, ImanAbuelMaalyAbdelrahman, "*Detection and Classification of Malaria in Thin Blood Slide Images*", 2017 International Conference onCommunication.

[11] https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria/

Fake Product Review Detection and Classification Using Data Mining

Dr. Rekha Sharma, Anamika Prasad,Manvi Saxena, Akshit Vaghani
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
rekha.sharma@thakureducation.org, anamikaprasad.98@gmail.com, saxena.manvi11@gmail.com, akshit.vaghani7@gmail.com

Abstract— **any e-commerce website gets a bad reputation if they sell a product that has poor reviews; the user most often blames the e-commerce website rather than the manufacturers. These online reviews can change the customer's mind set. If these reviews are true, then they can help users select the appropriate product to meet their need. On the other hand, it can confuse the user if the reviews are manipulated or not true.Some review sites include some great audits by individuals within the item organization itself to provide false positive item reviews. It also happens that rival businesses generate false negative reviews to outperform their competitors. To eliminate this type of fake product review, we will create a system that detects fake reviews and eliminates all fake reviews using data mining techniques. Finally sentiment analysis is performed to classify positive and negative for actual reviews**

Keywords— Sentiment analysis, Prediction, Data Extraction

I. INTRDUCTION

Now-a-days an internet has become an essential thing, as it provides more facilities to its users. There are many social networking sites that offer users to share their views. As the technology of e-commerce is growing rapidly, the impact of online reviews increases. For consumers, reviewing becomes an essential way for them to gain more knowledge of product quality thus enabling them to make purchasing decisions. For business owners, the review seeks to improve and grow their businesses. Feedback from customer reviews is useful for product improvement. However, reviews may not always be provided truthfully and fake reviews are usually present. Business owners can pay someone to write a good review about their products or a bad review about their competitors' products. These fake reviews lead consumers to make wrong decisions about product quality. Therefore, it is important to detect fake reviews and is a challenging issue in both the industrial and academic sectors. People share their views about politics, social issues as well as various products. It is common today that a user watches reviews of that product online before purchasing anything. There are many sites that deal with these reviews. They

Provide ratings for products as well as show comparisons between different products. Some enterprises try to create fake reviews to influence customer behavior and increase their sales. But, how to identify those fake reviews is a difficult task for the customers. In today's competitive world it is essential for any enterprise to maintain its reputation in one market. Therefore it is necessary for both, namely to identify the enterprise and customer to the manipulated reviews. Traditional methods of data analysis have long been used to detect fake reviews. A review can be classified as either fake or real using either supervised or unproven teaching techniques. In this work we try the dataset and determine whether the review is real or fake.

II. OBJECTIVES

The contributions of the presented work are as follows:

- To implement best approach available for detection of fake reviews using opinion mining (sentiment analysis) techniques.
- To detect and flag fake reviews posted on ecommerce websites.
- To let users, know if each individual review is trust worthy or not for efficient use of money from user's side.
- To tell whether it is a 'false positive or false negative'
- Review with highest accuracy.
- To make a provision of removal of such reviews.

III. METHODOLOGY

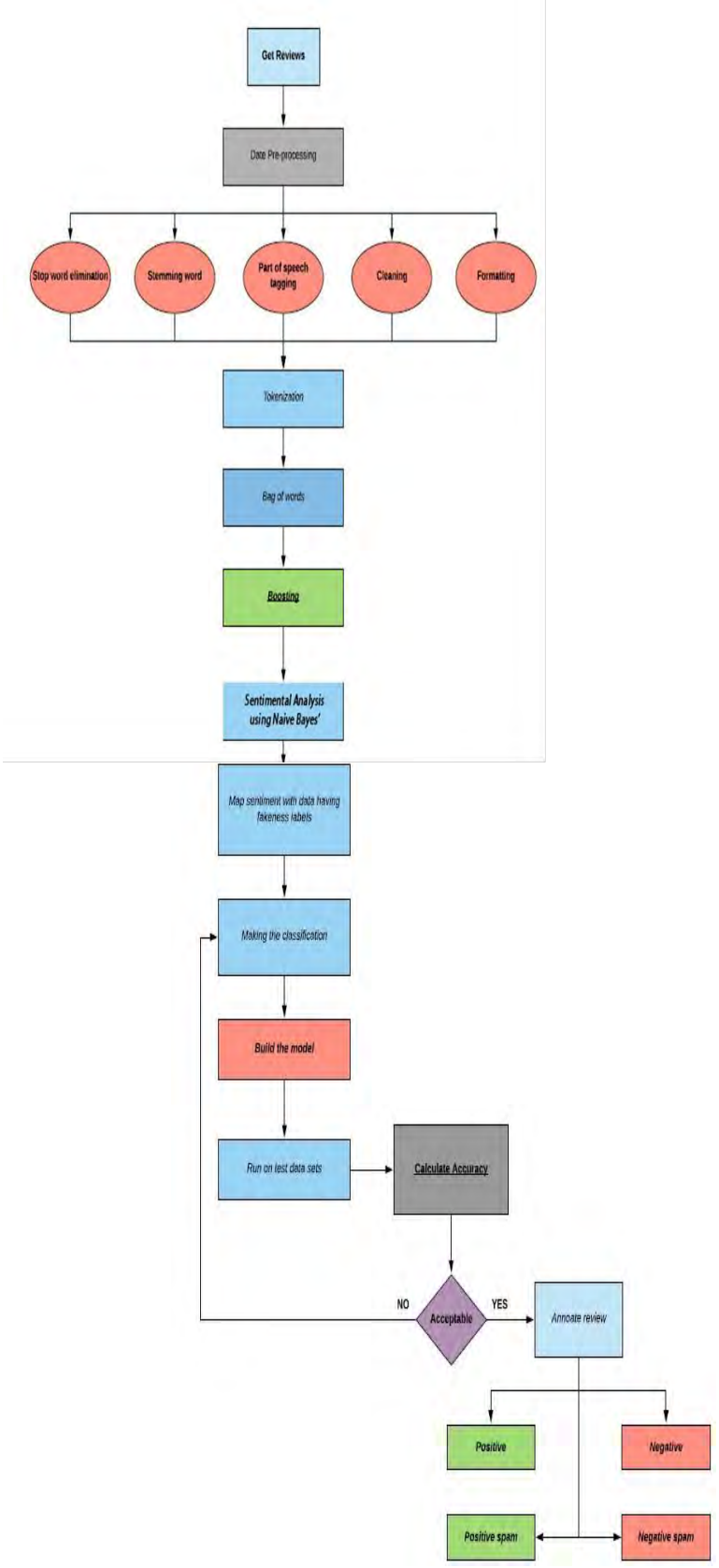
With our approach we aim to combine these 2 classifications and give a total of 4 classifications. The data preprocessing technique varies according to the dataset chosen. We use the methods like stop word elimination, part of speech tagging, tokenization. The use of algorithm also varies but all show varying accuracies. We have chosen Naïve Bayes' algorithm This is because it has better performance results. This approach is used in sentimental analysis so as to find whether the feedback is a positive or a negative one. Classification rules will be laid down, on the basis of the decision will be taken whether the review is fake or not.

Firstly, the most important aspect of the project is the data set. So the dataset is extracted by surfing over the internet. The dataset which best suits our interest is then selected. This dataset is then imported on the Machine Learning platform. This data is then distributed as 80% training and 20% test data. This data is the pre-processed using data cleaning methods like stemming word, stop word elimination, part of speech tagging etc. With this cleaned data, we train the classifier and then the

algorithm is applied. Naïve Bayes' algorithm is used for the sentimental analysis to achieve highest possible accuracy.

The proposed project work is under the Intelligent System Design and Development Domain which requires a detailed study of Natural language processing, Classification algorithms and Machine Learning.

The overall working of our project can be explained through the following abstract diagram.



A. Data Pre-processing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data are often incomplete, inconsistent and / or lacking in certain behaviors or trends, and are likely to include many errors. Data preprocessing ways include Stop word elimination, removal of URL, repeated characters, hash tags and usernames.

B. Training and Testing using Naïve Bayes algorithm

A Naïve Bayes classifier is an algorithm that uses Bayes' theorem to classify objects. Naïve Bayes classifiers assume strong, or naïve, independence between attributes of data points. Popular uses of naïve Bayes classifiers include spam filters, text analysis and medical diagnosis. These classifiers are widely used for machine learning because they are simple to implement.

C. Sentiment Analysis

Finally using Sentiment Analysis the result is obtained as positive or negative. The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. Natural language processing (NLP) is used for this purpose. It is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyse large amounts of natural language data. To implement natural language processing in python Natural language toolkit(nltk) was used. NLTK is one of the leading platforms for working with human language data and Python. In the following project we have used nltk to make predictions, that is , given a product review, a computer can predict if its positive or negative based on the text. Also, various packages are imported from nltk for the pre-processing step.

D . Determine Fake or not

Further, for the fake, genuine review detection, reviews will be processed and analyzed on the following conditions.

- Does the review entered by the user contain any link which redirects them to other product page for brand promotion?
- Analyzing whether multiple review have come from the same user.
- Analyses whether same email account or same IP

Address are used for multiple reviews on same product

- Analyze whether the review contains more than 5 positive or negative words
- Examine the timing of reviews. See if there is a spike in the total number of reviews during a very short time frame.
- Look for overuse of "I" and "me" and a lot of verbs. Online reviews that frequently use "I" and "me" are more likely to be fake than those that don't -- possibly because when people are lying they try to make themselves sound credible by using personal pronouns. If the review posted by the user

satisfies most of the above specified conditions then it will be considered as spam/fake reviews. Once the review is detected as spam review or fake review, then user account will be blocked and review will be discarded.

IV. CONCLUSION

It is seen that sentiment analysis play vital role to make business decision about product/services. Sentiment Analysis not only encompasses concepts of text mining but also the concepts of information retrieval. Major challenges in Sentiment Analysis includes feature weighting which plays a crucial role for good classification. Also it is seen that soft computing techniques have not been extensively used in the literature. Thus, with the blend of Tokenization, Decision tree algorithm and Sentiment analysis, higher accuracy of results in short duration can be expected. After investigating various machine learning techniques to build classification models using different features sets obtained Precision rates.

- Analyzing whether multiple review have come from the same user.
- Analyses whether same email account or same IP

Address are used for multiple reviews on same product

- Analyze whether the review contains more than 5 positive or negative words
- Examine the timing of reviews. See if there is a spike in the total number of reviews during a very short time frame.
- Look for overuse of "I" and "me" and a lot of verbs. Online reviews that frequently use "I" and "me" are more likely to be fake than those that don't -- possibly because when people are lying they try to make themselves sound credible by using personal pronouns. If the review posted by the user

Satisfies most of the above specified conditions then it will be considered as spam/fake reviews. Once the review is detected as spam review or fake review, then user account will be blocked and review will be discarded

Based on this research study, improvements can still be made. For example, more attributes can be

considered such as the time stamp, tracing backwards the user ID and IP address. Furthermore, more instances can also be collected. Although there are numerous machine learning models for sports results prediction, better models with high accuracy rates are still worth exploring especially function based techniques.

V. FUTURE SCOPE

The project uses analyzed historical data and tries to predict the outcome and classify them. The major benefit of the project is that the result will be shown on the spot. The main aim of the project will be to increase the accuracy of the result so as to enhance the performance as compared to the previous approaches .The project can be further used at a greater level in the industry by using boosting technique while pre- processing. Moreover, the project allows us to check the odds and false negative so that they get a clear idea as to which reviews should be considered to be worked upon.to be worked upon. Worked upon The project has a number applications. False positive.

VI. ADVANTAGES

The algorithm could also be used at any kind of platform with any kind of review as well if it works efficiently and with a lot of accuracy. It could also be used by any type of company owner to get to know their rivals in the market who are trying to defame them. Also it would help them understand the reviews better classified as positive and negative. Not only comments or reviews, the model will be able analyze any type of statement and classify it as positive and negative.

VII. ACKNOWLEDGEMENT

We sincerely thank our guide Dr. Rekha Sharma for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr.B. K. Mishra and the college management for their support.

References

[1] Long- Sheng Chen, Jui-Yu Lin, ”A study on Review Manipulation, Classification using Decision Tree", IEEE conference publication,2013Bunker, Rory & Thabtah, Fadi. (2017) “A Machine Learning Framework for Sport Result Prediction. Applied Computing and Informatics”, 15. 10.1016/j.aci.2017.09.005.

[2] Elshrif Elmurngi, Abdelouahed Gherbi, “An Empirical Study on Detecting Fake Reviews Using Machine Learning Techniques”, The Seventh International Conference on Innovative Computing Technology INTECH,2017

[3] Huaxun Deng, Linfeng Zhao, Ning Luo, “Semi-supervised Learning based Fake Review Detection”, IEEE International Conference on Ubiquitous Computing and Communications,2017

[4] Shweta Yadav, Deven Shah(2019), “Detecting Fraudulent Manipulation of Online Reviews”, International Research Journal of Engineering and Technology(IRJET) Volume 06 Issue 06, June 2019

[5] Piyush Jain, Karan Chheda, Mihir Jain, Prachiti Lade, “Fake Product Review Monitoring System”, International Journal of Trend in Scientific Research and Development (IJTSRD) Volume 3 Issue 3 Mar-Apr 2019

[6] Manali S.Patil, A.M.Bagade, “Online Review Spam Detection using Language Model and Feature Selection” International Journal of Computer Applications (0975 – 8887) Volume 59 No.7, December 2012.

[7] Kolli Shivagangadhar, Sagar H, Sohan Sathyan, Vanipriya C.H , “Fraud Detection in Online Reviews using Machine Learning Techniques “, International Journal of Computational Engineering Research (IJCER) May 2015 Engineering and Technology(IRJET) Volume 06 Issue 06, June 2019 [5] Piyush Jain, Karan Chheda, Mihir Jain, Prachiti Lade, “Fake Product Review Monitoring System”, International Journal of Trend in Scientific Research and Development (IJTSRD) Volume 3 Issue 3 Mar-Apr 2019

Design and Development of Enhanced Plagiarism Checker

Rahul Pal,Deepak Shukla,Atul Singh,,Dr. Rekha Sharma Associate Professor
Computer Engineering Department
Thakur College of engineering and Technology
Mumbai, India

Abstract— **The plagiarism of technical journals, papers and other documents is a widespread and growing problem in the academics as well as research process. The traditional manual detection of this kind of plagiarism by human is hefty, not accurate, and overloading process. This project aims to create a plagiarism-detection system that can help academicians and researchers to make a better judgment of the work. There are currently many existing detection algorithms. Some of them have been studied and analyzed and compared. Based on this, the most suitable for our system has been chosen for the implementation. The proposed system checks plagiarism between text documents. Both Syntactic and Semantic similarity is calculated and the similarity result will be shown stating whether two sentences are similaror not.**

Keywords — Plagiarism detection. Levenshtein Algorithm

I. INTRODUCTION

Martins et al. define plagiarism as "the usage of work without crediting its authors". The easy and cheap access to enormous web content has turned plagiarism into a serious problem be it for researchers, publishers or educational institutions. One of the most common forms of plagiarism in academia is the textual plagiarism at document level, where documents are typically essays or reports andscientificpapers.Accordingtoarecentsurvey,16perc entofthe original articles published in leading surgical journals could be considered redundant publication. However, plagiarism applies not only to text documents but also to source-code as well. Source-code plagiarism can be defined as trying to pass off (parts of) source code writtenbysomeoneelseasone'sown(i.e.,withoutindicati ngwhich parts are copied from which author). In academic environment, source-codeplagiarismariseinprogrammingassignments.With the intention of achieving good grades with less or almost no effort, students often try to copy the assignments from their friends. Freshmenwhoplagiarizeintheircoursesaremorelikelyt ocontinue this malpractice in their later courses. Therefore this malpractice needstobecurbedatitsinitialstages. Theinstructor ofacoursecan receive a false feedback about the level of the course and performance of the students. This makes the problem of assignment plagiarismdetectionanimportanttask.It ishardtomanuallyinspect and (decide whether a submission is genuine or plagiarized) detect similar student submitted solutions in a large class. Though manual inspection is effective, it is laborious and time consuming. One possible way to address this is to seek the help of automated code comparison tools like Measure of Software Similarity (MOSS), JPlagwhichhelpinidentifyingsimilarsubmissionpairs. Mostofthe well-knownautomaticcodecomparisontoolsrelysolelyonthe text- based approach or use the features solely based on the property of the assignments at a syntactic level to detect plagiarism. However, boththeseapproachessuccumb tocodeobfuscation,whichisahuge obstacle toautomatic software plagiarism detection. Often students use clever techniques to obfuscate the code and evade from being detected.

II. RELATEDWORK

[1] Bradley Beth,A Comparison of Similarity Techniques for Detecting Source Code Plagiarism, The University of Texas at Austin, May 12,2014.

This paper addresses techniques applicable to the phrase "...structure and computer code..." in relation to computer science coursework. Got to know different techniques of source-code attacks. There is room for improvement in implementing document fingerprinting techniques over source code

[2] Zhan Su, Byung Ahn, Ki-Yol Eom, Min-Koo Kang,Plagiarism Detection

Using the Levenshtein Distance and Smith-Waterman Algorithm, IEEE INSPEC Accession Number: 10184600, 22 August 2016.

The original aim to determine whether Levenshtein distance is a good enough value for plagiarism checking results in a positive, Levenshtein distance is a good enough value for plagiarism. Informally,

the Levenshtein distance between two words is the minimumnumberofsinglecharactereditsrequiredtochangeaword into another word. There exists strategies to fool the algorithm but the strategies can be avoided by using severalrestrictions.

[3] A Machine Learning Based Tool for Source Code Plagiarism Detection, Upul Bandara, and GaminiWijayarathna,2014.

It gives several attributes to measure code similarity. This paper presents a new plagiarism detection method, which is based on machine learning techniques. We have trained and tested machine learning algorithms for detecting source code plagiarism. Most system will not work correctly, if programmers follow some coding standards and source code formatting tools specified in their projects.

III. PROBLEM DEFINITION

The objective of this project is to make a plagiarism checker software or tool which checks plagiarism between documents or betweendifferentsourcecodes.Andatlastthesimilarity percentage is computed with respect to syntactic and semantic similarities for documents and line wise checking for source codes and thus giving results whether the document is plagiarizedornot.

Inthiswork,ourfocusisonbuildingatoolpoweredbymachine learning algorithms, to detect plagiarism in programming assignments. We propose and use novel features that can alleviate the effect of code obfuscation and be able to detect the plagiarized student solutions. We use source code metrics as feature representations which can capture the variations observed in the code as belonging to three distinct themes: structure, syntax (syntax refer to the tokens that occur within basic blocks) andpresentation.

$$lev(i,j) = \begin{cases} \max(i,j), \min(i,j) = 0 \\ \min\begin{cases} lev(i-1,j)+1 \\ lev(i,j-1)+1 \\ lev(i-1,j-1)+eq(i,j) \end{cases}, \min(i,j) \neq 0 \end{cases}$$

IV. OBJECTIVES

- Detection Performance: The detection performance should be better than that of other tools, and we should be able to demonstrate this on a variety oftasks.

- Visualization: To be usable as plagiarism detection tool, it is important to show the locations of the similarities in each programpairorthesourceofotherfeatures.
Weshouldbeable to visualize the influence of certain features in the document pairs to make comparingeasier.

- Runtime Performance: Finding plagiarism in a set of source code documents should not take days, but preferably less than an hour. Besides the inconvenience for the user, slow performance makes evaluation of the tool and hyper parameter optimization harder. However, improvements in detection performance may justify a higherruntime.

V. METHODOLOGY

- *LevenshteinDistance*

Levenshtein distance is a measure of difference between two strings. Informally, the Levenshtein distance between two wordsistheminimumnumberofsinglecharactereditsrequired tochangeawordintoanotherword.Themeasurementwas first considered by Vladimir Levenshtein, and his name is used as the distance name. The distance may also be referred as edit distance, although the edit distance term actually refers to a largersetofstringdifferencemeasure,suchaslongestcommon subsequence (LCS) and Hamming distance. Formally defined, given two strings and its character set, the edit distance is the minimum-weight series of edit operations that transforms the first string into the second string. One of the simplest sets of editoperationsdefinedbyLevenshteinin1966are:insertion of a single symbol, deletion of a single symbol, and substitution of a single symbol. The original definition has a unit cost, so the Levenshtein distance is equal to the minimum number of steps to change the first string into the second string. This definition is referred as Levenshtein distance. The levenshtein distance between two strings a,b is given as lev(length(a),length(b))

Figure 1: Levenshtein Distance

VI. FLOWCHART

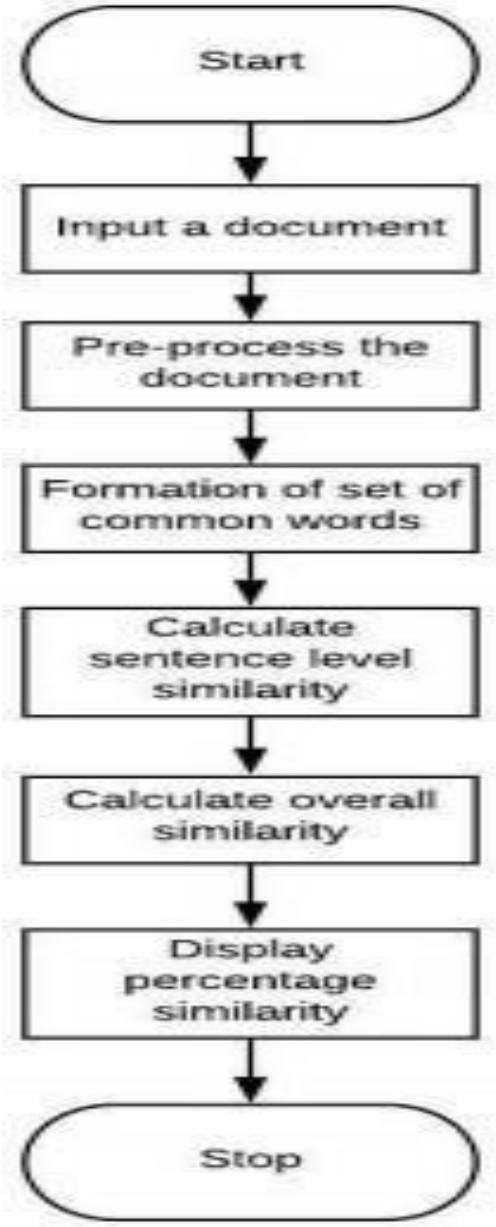


Figure 2: Flowchart

VII. BLOCKDIAGRAM

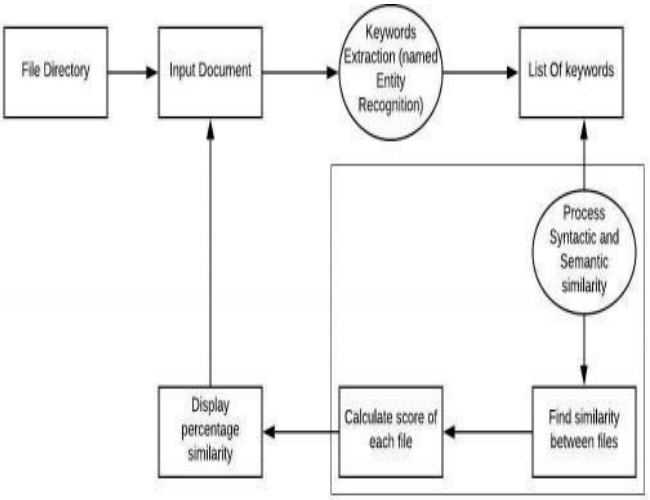


Figure 3: Block Diagram

VIII. CONCLUSION

Project of plagiarism checking using various Algorithms, systems such as MOSS and Machine Learning aim to check the plagiarized content for the requested document. Since the field of researchanddevelopmentisbloominganditrequesttheu rgencyfor the system that validates the uniqueness of the technical journals, paper and many other documentations. There are already existing system that compare and evaluate the plagiarism of a particular document but that comparison is done only on the basis of the syntactic similarities and judging the document to be plagiarized onlyforsyntacticsimilaritiesisnotafairplatform.

Inthefirstphase of our project we have partially implemented our proposed work on sentence similarities level basis. Wherein, we have computed syntactic and semantic similarities between two sentence and display theresult.

IX. ACKNOWLEDGMENT

WewishtoexpresssincerethankstoourComputerDepartment forarrangingsuchaniceevent,sothatwecanresearchallthisabo ut plagiarism detection methods. We also extend our heartfelt thanks to our colleagues, family members, andwell-wishers.

References

[1] Zhan Su, ByungAhn, Ki-YolEom, Min-Koo

Kang,Plagiarism Detection Using the Levenshtein Distance and Smith-Waterman Algorithm, IEEE INSPEC Accession Number: 10184600, 22 August 2016

[2] Tapan G., Hiren Joshi, Hardik Joshi, Source Code Plagiarism Detection, International Journal of Computer Applications (0975 – 8887) Volume 105 – No. 17, November2018

[3] Bradley Beth,A Comparison of Similarity Techniques for Detecting Source Code Plagiarism, The University of Texas at Austin, May 12, 2017.

[4] GeorginaCosmaandMikeJoy.Anapproachtosourcecodeplagiarism detection and investigation using latent semantic analysis. Computers, IEEE Transactions on, 61(3):379– 394,2016.

[5] Alberto Barrón-Cedeño and Paolo Rosso. On automatic plagiarism detection based on n- grams comparison. In Advances in Information Retrieval, pages 696– 700. Springer,2014.

[6] Zoran Đurić and Dragan Gašević. A source code similarity system for plagiarism detection. The Computer Journal, 56(1):70–86,2016.

Flight delay prediction using machine learning algortihms

Dhanshree Kolhe, Dolby Agarwal, Anilkumar Yadav
Department of Computer Engineering
Thakur college of Engineering and Technology
Mumbai, India
kodhanhsree@gmail.com, dolby14agarwal@gmail.com, anilyadav181997@gmail.com

Abstract—***The goal of the model proposed in this paper is to predict airline delays caused due to the major observed factors***
i.e. weather conditions, NAS(National aviation system), Airline and airport security using data mining and supervised machine learning algorithms. US domestic flight data and airline delay data from 2005 to 2015 were extracted and used to train the model. To overcome the effects of imbalanced training data, sampling techniques are applied. Decision trees, random forest and the XGBoost were implemented to build models which can predict delays of individual flights. Then, each of the algorithms' prediction accuracy and the receiver operating characteristic (ROC) curve were compared. In the prediction step, flight schedule and delay details were gathered and fed into the model. Using those data, the trained model performed a binary classification to predicted whether a scheduled flight will be delayed or on-time.

I. INTRODUCTION

According to the 'Bureau of Transportation Statistics (BTS)', approximately twenty percent of the entire scheduled commercial flights are delayed. Due to airline delays, airlines are costed multi-billion dollars per year and cause inconvenience to passengers. BTS has categorized airline delays into five main causes, which are air carrier, extreme weather, National Aviation System, late-arriving aircraft and security. All the factors mentioned above are closely related to other categories, indeed. For example, National Aviation System category can include delays due to the re-routing of flights by inclement weather. Besides, weather too is a factor affecting late-arriving aircraft although airlines won't report the causes as weather. By considering those facts, weather's percentage share accounts for about 40% of total delay minutes. Thus, study on the influence of inclement weather on airline delays is essential for efficient flight operations in relation with all the other factors. Furthermore, a decision support tool built on the study can inform passengers and airlines about delays in advance and help them reduce possible monetary losses. For this purpose, the classification model to predict delays of individual flights is proposed in this study to follow. On- time performance of flights has been an important research subject since the demand for air travel is ever increasing. Thus, several attempts were there to discover patterns in air traffic. Rebello and Balakrishnan have created variables indicative of the state of the NAS. And they predicted network-related delays of the future by studying the system- level dependencies among airports. Zonglei et al. have trained classification models to find how serious daily delays are at the hub-airport of China in the future. Contrastingly, Klein et al. have focused more on weather and provided a delay prediction model established on metric called Weather Impacted Traffic Index (WIFI) which measures the severity and the impact of weather. However little research exists that focuses on combined factors analysis of delay prediction of individual flight by utilizing machine learning. In this research, we have focused on arrival delays of individual flights using supervised machine learning algorithms. There are various reasons to explain why machine learning was tried. First of all, the volume of historical flight and delay data are too large to analyze analytically. Moreover, relationships between causal factors and delay or even correlations among factors are very complicated and highly nonlinear to test all hypothesis. Machine learning is able to develop models vigorously with huge amount of dataset and it has the ability to discover hidden patterns in the data. To summarize, machine learning is a clever method that can address problems in analytical analysis with big data. For predicting delays of individual flights, supervised machine learning algorithms were implemented with features including flight schedules and delay data at the origin and the destination airport. In order to increase the predictive capability, models were trained for individual origin-destination (OD) pair not for the entire National Airspace System (NAS) by capturing each airport's delay data. In the prediction step, flight schedule information combined with delay data was fed into the model to get the predictions on scheduled flights not yet flown. The paper is organized as follows: Section II provides brief explanation on methodologies used in this study and Section III describes how the model is built. In Section IV, classifiers' performance are investigated in terms of accuracy and ROC. Lastly, conclusions and future work are presented in Section V and they conclude the paper.

II. LITERATURE REVIEW

From the literature we reviewed, we found that this problem of flight delay prediction had different approaches by different people. The flight delay

prediction problem may be modeled by many methods, depending on the objectives of the researches. In this project we have approached the problem from machine learning perspective. Researches that study flight systems from machine learning perspective are

increasing. The methods commonly used include k-Nearest Neighbour, neural networks, SVM, fuzzy logic, and random forests. They were mainly used for classification and prediction.

Rebollo et al. applied random forests to predict delay innovation. They compared their approach with regression models to predict delay innovation in airports of the United States considering time horizons of 2, 4, 6 and 24 hours. Lu et al. compared the performance of Naïve Bayes, decision tree and Neural networks to predict delays in large datasets. They observed that decision trees had the best performance with confidence of 80%.

Lu et al. built a recommendation system to forecast delays at some airports due to propagation effects. Prediction was based on the k-Nearest Neighbour algorithm and used historical data to recognize similar situations in the past. The authors noticed fast response time and easy logical comprehension as the main advantages of their method. Khanmohammadi et al. created an adaptive network based on fuzzy inference system to predict delay innovations. The predictions were used as an input for a fuzzy decision making method to sequence arrivals at an International Airport in New York. of classifiers is dependent on the previous one and focuses on the previous one's errors. Samples that are incorrectly predicted in the previous classifiers are chosen more often or weighted more heavily when estimating a new classifier.

III. MOTIVATION

Flight delays have negative impacts, especially economic, for passengers, airlines and airports. Given the uncertainty of their occurrence, passengers usually plan to travel many hours before their appointments, increasing their trip costs, to ensure their arrival on time. On the other hand, airlines suffer penalties, fines and additional operation costs, such as crew and aircrafts retentions in airports. Furthermore, from the sustainability point of view, delays can also cause environmental damage by increasing fuel consumption and gas emissions. There is a known relationship between levels of delays and fares, aircraft sizes, flight frequency and complains about airline service. The estimation of flight delays can improve the tactical and operational decisions of airports and airlines managers and warn passengers, so they can conveniently rearrange their plans. It will be good experience to work with real data and practice skills related to statistics and machine learning. Also, it is interesting to discover more insights for airline industry.

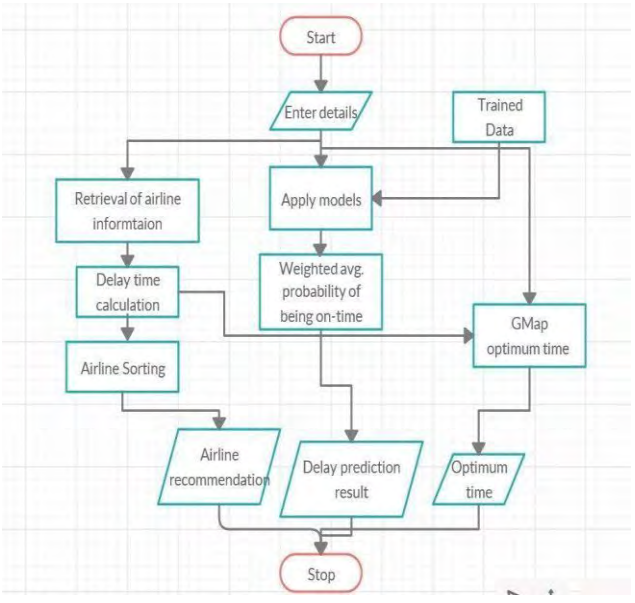
IV. METHODOLOGY

A. Supervised Learning Classifiers

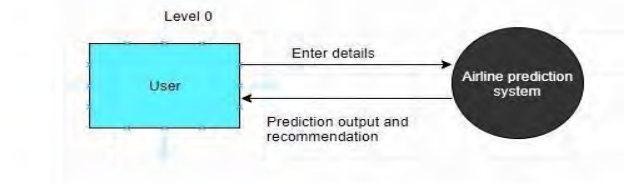
Random Forest (RF): Random forest is an ensemble of many individual decision trees. It builds a large collection of de-correlated trees which are noisy but unbiased, and averages them to reduce the variance. Random forest obtains a class vote from each tree, and then classifies a sample using majority vote. Let Cˆb(x) be the class prediction of the bth tree, then the class obtained from random forest, Cˆrf (x), is Cˆrf (x) = majority vote{Cˆb(x)}B

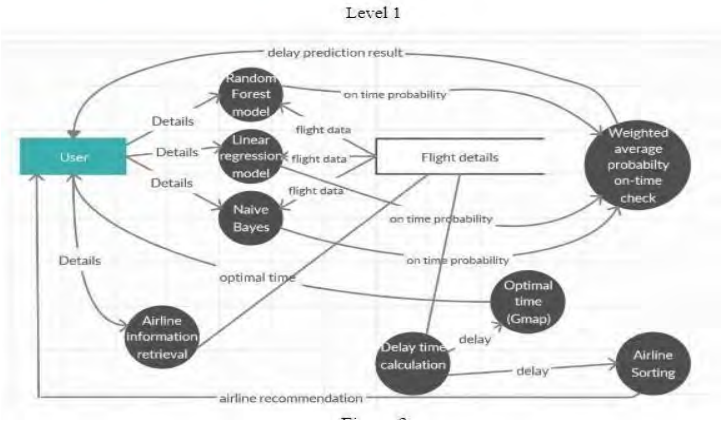
XGBoost: AdaBoost is a method of converting weak classifier into highly accurate prediction rule. It learns week learner on weighted example set sequentially and combines weak hypotheses linearly. Produced sequence

B. Flowchart

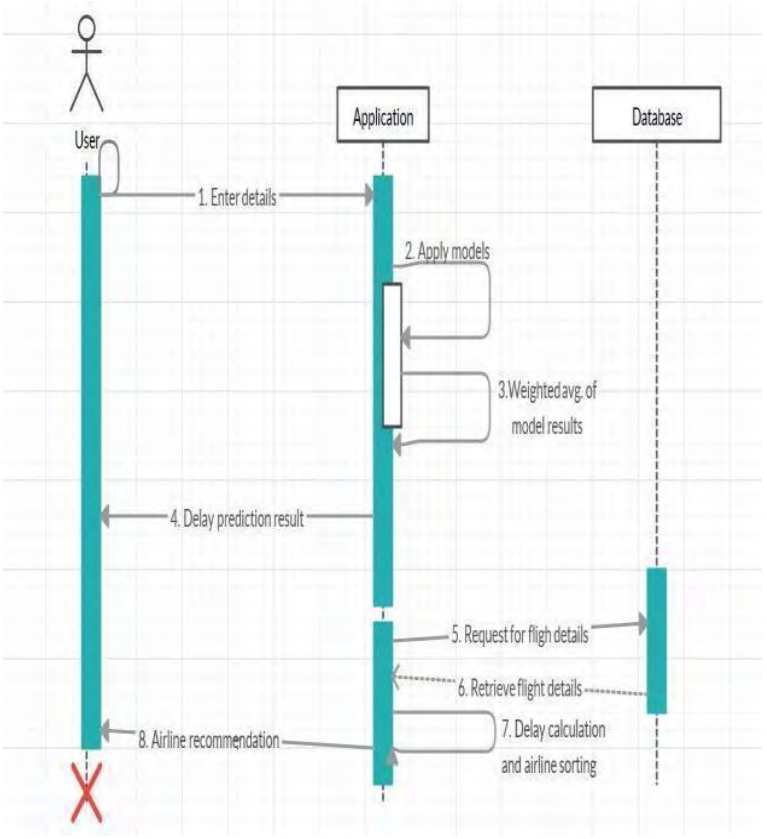


C. *Functional Model*





D. *Sequence Diagram*



V. APPLICATION OF PROJECT

Optimum Time Management: The proposed system is giving the percentage of the possibility of the flight getting delayed. The same system can be used to provide a specific time range where the possibility of the flight getting delayed is very low by considering the same attributes, the user can book their flight in that time range.

Travel Insurance: The system can be used by the Insurance Companies to get an estimate of the amount that will be used for settling the claims in advance. The claims can include travel delay, missing a connecting flight, baggage missing, etc.Hotel Management: The system can be used by the Hotel to check if a flight is delayed or not and sending the vehicles for pickup and drops accordingly. The hotel can obtain the itinerary details of the customers to schedule thepickup and drops.

VI. RESULT AND ANALYSIS

A. Result

We were able to use the random forest classifier to successfully classify the dataset that we had. We applied data preprocessing and cleaning techniques using Python to remove the unnecessary attributes. The
accuracy was 86% using Random Forest classifier.

B. Futher Work

Further work needs to be done so as to apply various other algorithms and take a weighted average of their accuracies since different algorithms work best on different datasets.

References

[1] Prediction of Weather-induced Airline Delays Based on Machine Learning Algorithms Sun Choi, Young Jin Kim, Simon Briceno and Dimitri Mavris Aerospace Systems Design Laboratory Georgia Institute of Technology Atlanta, Georgia 30332–0250 - IEEE

A Statistical Approach to Predict Flight Delay Using Gradient Boosted Decision Tree Suvojit Manna, Sanket Biswas, Riyanka Kundu Somnath Rakshit, Priti Gupta and Subhas Barman Department of Computer Science and Engineering Jalpaiguri Government Engineering College, West Bengal, India. 2017 International Conference on Computational Intelligence in Data Science(ICCIDS)- IEEE

[2] Airline delay prediction by machine learning algorithms H. Khaksar and A. Sheikholeslami Department of Transportation Engineering and Planning, School of Civil Engineering, Iran University of Science & Technology, Tehran, Iran. Received 24 May 2017; received in revised form 15 October 2017; accepted 23 December 2017.

[3] U. D. of Transportation, “February 2016 on-time performance up from previous year,” jan 2016.

[4] H. B. Juan Jose Rebollo, “Characterization and prediction of air traff ic delays,” Transportation Research Part C: Emerging Technologies, vol. 44, pp. 231–241, 2014.

[5] Z. G. Lu Zonglei, Wang Jiandong, “A new method to alarm scale of f light delay based on machine learning,” in Proceedings of International Symposium on Knowledge Acquisition and Modeling, 2008..

[6] R. S. L. Alexander Klein, Chad Craun, “Airport delay prediction using weather-impacted traffic index (witi) model,” in Preceedings of the Digital Avionics Systems Conference(DASC) 29th, 2010.

[7] L. O. H. W. P. K. Nitesh V. Chawla, Kevin W. Bowyer, “Smote: Synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

Implementation Of A System For Real Time Color Recognition

Rajan Jha, Hasrh Jain, Vaishnav Baswe, Vidyadhari Singh
Computer Engineering
Thakur College of Engineering and Technology
Indiarajanjha.work@gmail.com, Indiaharshashokjain1@gmail.com , Indiavaishnavbaswe@gmail.com, vidyadhari.rks@thakureducation.org

Abstract— ***Accurate color recognition models are a requirement for the growth of technology when concerned in dealing with Artificial Intelligence as well as in many industries whose interest lies in automated segregation of colors such as the publishing industry. In light of numerous such requirements and with the information of complexion of various images, machine learning calculations based on Manhattan distance can recognize the similarity of a new image as compared to the assigned dataset and characterize it into a particulargroup.***

There has been an extensive use of a color recognition model which can help various industries in streamlining their processes and reducing cost. Color recognition has a very wide uses ranging from science and technology to society. The project mainly focuses on building a system which can learn to classify images based on their colors. The system will learn from the dataset to classify new attributes using Machine Learning. The system will be focusing on effectively dealing with primary colors and will classify them based on majority color constituents of the image. The system evaluates the color histogram of the image and learns from the data set to classify it accurately. Integration of such a system with webcam can dynamically help the AI industry in achieving its goal.

Keywords— Histogram, OpenCV, Python, Color Recognition

I. INTRODUCTION

This undertaking centers around color recognition by K- Nearest Neighbors Machine Learning Classifier which is prepared by R, G, B Color Histogram. It can group Red, Green, Blue, White, Black, Violet, Orange, and Yellow colors. On the off chance that we need to characterize more colors or improve the precision for some critical applications we should chip away at the training data set or consider about other color highlights, for example, Color Moments or Color Correlogram.

We disengage the assignment into three obvious stages**:**

- **Training K-Nearest Neighbors Classifier:** Train KNN classifier by R, G, B Color Histogram esteems.

- **Feature Extraction:** Perform feature extraction for getting the R, G, B Color Histogram esteems of preparing pictures for trainingimages.

- **Classifying via Trained KNN:** Read Web Cam outline by outline, perform feature extraction on each edge and afterward arrange its mean shade via prepared KNNclassifier.

II. METHODOLOGY

The 2 main phenomena to understand basic Object Detection/Recognition Systems of Computer Vision and Machine Learning are:

1) Feature Extraction

We focus on a major to represent the points which seem of interest and then compare them with other such similar points in the image.

2) Classification

An algorithm that actualizes grouping, particularly in a solid implementation, is known as a classifier. The expression "classifier" now and then likewise alludes to the numerical function, executed by a grouping algorithm that maps input information to a classification.

Color Histogram

Color Histogram is a portrayal of the dissemination of colors in a picture. For advanced pictures, such as digital images a color histogram speaks to the quantity of pixels that have colors in every one of a fixed list of color ranges, that span across the picture's shading space, the arrangement of every single imaginable color. Color Histogram is the most broadly utilized method for extricating the color feature of a picture. It represents the picture from an alternate point of view. It represents the recurrence appropriation of shading canisters in an image. Itcheckscomparable pixels and store it. There are two sorts of color histogram viz. global color histogram and local color histogram. Color histogram is proposed as a worldwide color descriptor which break down each measurable color frequency in a picture. It is utilized to take care of the issues like change in translation, pivot and point of view. Local color histogram centers around the individual pieces of an image. Local color histograms consider the spatial distribution of pixel which is lost in global color histograms. Color histogram is anything but difficult to register towards little varieties in the picture so is significant in ordering and recovery of picture database. Aside from these focal points, it faces two significant disadvantages. To start with, overall and large spatial data isn't considered. Second

is that the histogram isn't powerful and extraordinary as two unique pictures with comparative color appropriation lead to comparable histograms, a similar view pictures with various presentation to light make varioushistograms.

Color Histogram for K NearestNeighbours (CHKNN)

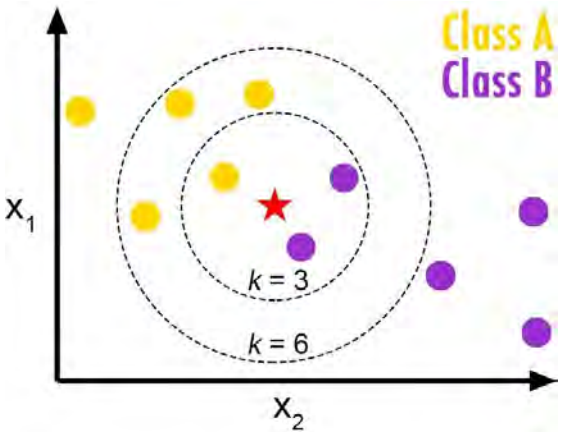
For a color pixel 2^{24} different colors are possible. The color histogram for K-Nearest neighbor (CHKNN) features of an image is presented. In this method one color in the common color palette which is more similar to a particular pixel color is considered to replace that pixel color. Same process was applied to each pixel color in order to classify all pixels in the image into k cluster. The K-Nearest Neighbor clustering algorithm is employed to categorize the pixels of all training images. The algorithm outputs the mean of all pixels in each cluster, which is set as the new initial value for further training. After these steps the color histogram is prepared. For kth color bin the CHKNN feature was defined as shown in Eq:

$$g_k = \frac{N_k}{N}$$

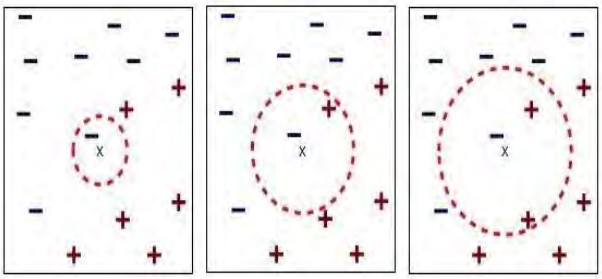
Where, N is total pixel numbers and N_k is the number of the pixels in the kth cluster. The method describes color feature with a small number of features and hence 16 groups of colors were evenly divided in the paper leading to effectively shorten image retrieval time and enhance retrieval performance. It is simple, easy to compute and noise resistant method. The method is robust to image size variance, displacement androtation.

K-Nearest Neighbors Algorithm

The k-nearest neighbors (KNN) algorithm is a straightforward, simple to-execute supervised ML algorithm that can be utilized to solve both classification and regression problems. K nearest neighbors is an algorithm that stores every single accessible case and groups new cases dependent on a likeness or similarity measure, for example, a separation function like Euclidean distance. K-nearest neighbor has been utilized in factual estimation and pattern recognition as of now as a non-parametric strategy. The K-nearest neighbor calculation accept that comparative things exist in closeness. At the end of the day, comparable things are close to eachother.



The thought behind k-Nearest Neighbor algorithm is



(a) 1-nearest neighbor (b) 2-nearest neighbor (c) 3-nearest neighbor

K-nearest neighbors of a record x are data points that have the k smallest distance to x

very direct. To group a new record, the framework finds the k nearest neighbors among the preparation or training archives, and uses the classes of the k nearest neighbors to weight the classification competitors. One of the downsides of KNN algorithm is its efficiency, as it needs to contrast a test report with all samples in the training set. Furthermore, the efficiency of this algorithm relies significantly upon two factors, that is, a reasonable similitude work, in other words a similarity function and a proper value for the parameter k in K nearestneighbor.

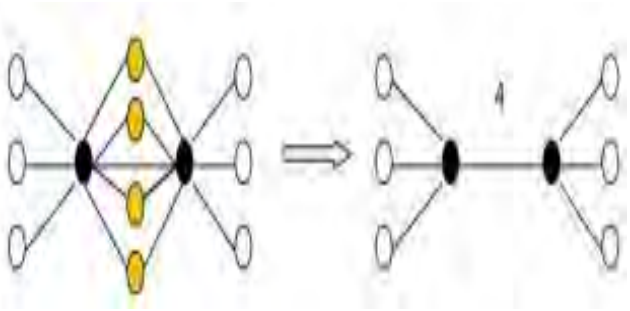
Improved k-Nearest Neighbor Classification algorithm

According to paper proposed by LeventErtoz, shared nearest neighbor method named SNN can be used to solve the problem of high-dimensional data which can be used in rapid recognition models such as real time color recognition models. The SNN core idea is that the similarity between two objects is defined based on Shared nearest neighbour. It is based on the following principle, if two points are similar to some of the same point, they are also similar, even directly similarity measure can't point out. The computation of SNN similarity between two point can be described as follows and Graphic is shown infigure.

Step 1: For each point, find out N nearest points for it and create its nearest neighbour list.

Step 2: If both two points appear in each other's nearest neighbour list, there is a link between them.

We build apoint link graph after we complete calculating link strength of every pair of linked points.
Step 3: The link strength of every pair of linked points is the number of shared neighbors.



Consider the above figure, each of the two black spots have eight nearest neighbors' interaction. The four of these nearest neighbors represented in yellow are Shared by the black dots. So, the SNN similarity between the two pointsis
4. Because SNN similarity is good on determining similarity between point and point, according to the SNN similarity we can successfully improve KNN classification algorithm to make it suitable for rapid recognition models.

3) PROPOSED WORK

We will likely recognize the color in still and dynamic images. We have thought of training the algorithm on several general color after which we read web cam frame by frame, perform feature extraction on each frame and then classify the mean color of it by trained K-Nearest Neighbor classifier. The objective is to first perform feature extraction for getting the R, G, B color Histogram of training images. The training images are mainly Black, Blue, Green, Orange, read, violet, white and yellow along with some other colors that are to be incorporated in the future. The goal is to set a connection between training images and new images captured through Webcam and color histogram seems to be the best guide.
The system tries to group the pixels based on the Euclidean distance. The Euclidean distance between any two points p and q is given by the length of the straight line connecting them. In Cartesian coordinates system if p having a space of points ($p1$, $p2$,..., pn) and q having a space of points ($q1$, $q2$,..., qn) are two points in Euclidean n-space, then the distance represented by d between p and q is given by the Pythagoreanformula

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n -$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

After clustering we discover the color in some group of pixels and again perform the clustering on these groups to ultimately achieve the color of the image. For Real time

Web cam recognition each frame is considered as an image and is recognized independently.

4) CONCLUSION

With growing use of color recognition models such as in the distribution industry like publication media to reduce the cost of printing by segregating the images possessing majority components of similar color distribution, an accurate and efficient recognition model has to be developed. There is no denial that such a model can be widely used in monitoring changes and in advance fields like Artificial Intelligence. Though there are some tools available in the market, many of them are not accurate and stable. The main difference lies in those tools using a focal point to recognize the color, whereas our model considers the entire image frame and finds the color having the maximum distribution in the image. To give an edge to our recognition model, our project aims to also take into consideration some important factors like accuracy, stability, less response time, large variety of colors. All these important factors along with large training dataset assures that our model will be able to provide better results. It is very easy to integrate the model with different applications.
We have optimized the KNN algorithm as per our requirement of rapid color recognition to achieve better efficiency and accuracy. In this report, an intelligent system was developed based on machine learning and feature extraction to deal with the problem of color recognition in images. After investigating various machine learning algorithms to build recognition models using different features set obtained by feature extraction methods, we arrived at a conclusion of using K-Nearest Neighbor with improvement based on SNN (Similar Nearest Neighbor). There are still many factors affecting the recognition of color. First and foremost is the quality of training data. Other features include orientation, size and varying exposure of light. Theoretically, coverage of more features will result in a better accuracymodel.
Various improvements have been made for the system to support real time color recognition with less response time. For example, Large datasets are used carrying information of various colors. Although there are numerous machine learning models for recognition, better models with high accuracy are still worth exploring especially function based techniques. The system was successful in proving itself as a color recognition model and was able to recognize maximum color distribution in different images.

REFERENCES

[1] Huang, J. and Zabih, R. (2019). *Combining Color and Spatial Information for Content*. [online] Cs.cornell.edu. Available at: http://www.cs.cornell.edu/rdz/Papers/ecdl2/spatial.htm.

[2] Senthamaraikannan, D. Shriram, S. and William, D. (2019). *REAL TIME COLOR RECOGNITION*. [online] Ijireeice.com. Available at:https://ijireeice.com/wp-content/uploads/2013/03/IJIREEICE3C-s- senthamaraikannan-real-time.pdf.

[3] M, R. and Niraimathi, D. (2019). *COLOR BASED IMAGE SEGMENTATION USING CLASSIFICATION OF K-NN WITH CONTOUR ANALYSIS METHOD*. [online] Irjet.net. Available at: https://www.irjet.net/archives/V3/i10/IRJET-V3I10222.pdf.

[4] Augustine, D., Jose, S. and T N, N. (2019). *Comparitive Study On Color Recognition Methods*. [online] Ijcstjournal.org. Available at: http://www.ijcstjournal.org/volume-3/issue-2/IJCST-V3I2P12.pdf.

[5] Zheng, W., Wang, H. and Ma, L. (2019). *An Improved k-Nearest Neighbor Classification Algorithm Using Shared Nearest Neighbor Similarity*. [online] Pdfs.semanticscholar.org. Available at: https://pdfs.semanticscholar.org/87f0/e0e4bf0332f0c81adf5b8379f9cc00c10d8e.pdf.

[6] Puri, R., Gupta, A. and Sikri, M. (2019). *CONTOUR, SHAPE, AND COLOR DETECTION USING OPEN CV – PYTHON*. [online] Iraj.in. Available at: http://www.iraj.in/journal/journal_file/journal_pdf/12-451-152482950826-27.pdf.

[7] Prabhakar, A., Devi, N. and Devi,R. (2019). *DIFFERENT COLOR DETECTION IN AN RGB IMAGE*. [online] Journalijdr.com. Available at: https://www.journalijdr.com/sites/default/files/issue- pdf/9748.pdf.

[8] Goel, V., Singhal, S. and Jain, T. (2019). *Specific Color Detection in Images using RGB Modelling in MATLAB*. [online] Ijcaonline.org. Available at: https://www.ijcaonline.org/archives/volume161/number8/go el-2017-ijca-913254.pdf.

Review Paper On Image Forgery Detection

Lokesh Agrawal Heth Jain Dhruv Agrawal

Abstract **The arrival of sophisticated image editing software has made manipulating digital images' child's play. Today visual inspection can rarely distinguish between an authentic image and a forged one. Our confidence in the integrity of an imagehasbeenconstantlydwindling.Frommain-streammedia to various social media platforms, political campaigns, and the mass forwarded photos on instant messaging services, manipulated photos are on the rise. Revolutionary techniques have been devised to counterfeit the negative effects of these forged images. Here we review some of thesetechniques.**

KEYWORDS:-Gradient, image forgery, detectors, deep networks, copy- move

I. INTRODUCTION

For many years, human beings have used images or pictures to capture the reality around them or modify it to send the message they want to convey. This evolution has a before and after with the creation of photography. In today's age, there are a significant amount of computer crimes related to the illegal possession, distribution or modification of multimedia content. However,manipulationofimageshasbeengoingonforvarious years and is not exclusive to the current age. It was possible to manipulate through splices the negatives of the photographs, for example in Figure 1 the famous photo of Soviet dictator Iósif Stalin along with his commissioner for Internal Affairs, who disappears from the photo by order of Stalin after being executed in1940.

In earlier times, manipulation of images had a political, re- ligious, cultural motive but now this activity is mainly carried out for advertising or aesthetics. The ease of manipulating the images has become easy due to technological advances in mobile phones, computers, photo editing applications like Adobe Photoshop, Adobe Premiere, GIMP etc. Nowadays, image can also be manipulated on mobile phones using applications like Picsart , Snapseed ,etc. Due to the advent of Social Media, distribution of fake or manipulated images have become far more easier than ever.

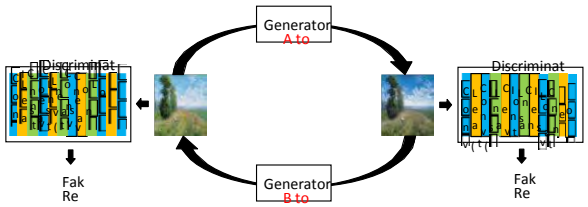


Figure 1: Cycle-GAN based image-to-image translation

InJuly2017,researchersfromthejournalCognitive Research
[2]usedadatasetof40scenes,30ofwhichweresubjectedto five different types of manipulation, including physi- callyplausible and implausible manipulations.707participants were showntoassesspeople'sabilitytodetectrealworldmanipulated scenes. The study found that only 60% of the people were able to detect the fake scenes and even then, only 45% of them were able to tell exactly where the altered content was. One of the main conclusions from their study was that if people cannot differentiate between real andfalsedetails in scenes, manipulations could often modify what we believe and remember.

Thereareanumberofimagetamperingtechniqueswhichare present which are used to manipulate or tamper the image. There are some valid reasons for tampering images such as marketing, promotional activities, etc. Unfortunately, the tools developed for these tasks are being used by some people to tamper the truth. In this paper, we take a look at some of the techniques which are used for detection of fakeimages.

This paper presents four algorithms; the first one is anError Level Analysis (ELA) algorithm, which can be used as an initial filter to detect the presence of splicing in an image. The second algorithm is a digital image authentication method based on the quadratic mean error of the Colour Filter Array (CFA)interpolationpatternestimatedfromtheanalyzedimage.
Thethirdalgorithmisaregionduplicationdetectionmethodfor images, which utilizes statistical image features in the frequency domain, viz., mean and variance. The primary motivation is to exploit the local

statistics of an image, using block-based mean and variance computations. The last algorithmusesanumberofdetectorstodetectfakeortampered images which are generated using Image to Image translation, based on generative adversarial networks (GANs). This manipulation technique is very dangerous as it allows one to modifycontextandsemanticsofimagesinaveryrealistic way. Therefore, we use methods based on CNN architecture, some based on ImageNet, etc. which we look further in thepaper.

*a) **1) GAN:-**:* As we know that image forgery and image manipulation is a big problem right now and due to advent of toolssuchasphotoshopeventhenon-expertuserscandoimage manipulation. There are many techniques for image manipulation but one of the most dangerous one is Image to imagetranslationusingGAN.Wewillreviewhowdetectors for imageforgerycanbeusedfordetectionoffakeimagesand how we can overcome this problem testing the cases in both ideal and compressedconditions.

Computer fares better when it comes to detect fake images which are not computer generated. Study shows that 62- 66% oftheimageswereclassifiedandtheusersfared worsethanthe computer. But when it comes to computer generated fake images, the trend is reversed. This may be attributed to the inability of current computer graphics tools to provide a good levelofphotorealism.Somepapersproposeto detectcomputergraphics images based on statistics extracted from their wavelet decomposition or from residual images while some papers rely on the different noise introduced by the recording device on traces of chromatic aberrations or traces of demosaicing filters. In face asymmetry is proposed as a discriminative feature to tell apart computer generated from natural faces. Only very recently, deep learning has been used for this task, and found to outperform preceding approaches. We analyze the performance of a number of learning- based methods in the detection of image-to-image translation. Our goal is to understand if, to which extent, and in which conditions, these attacks can be unveiled. Also to solve this problem, state-of-the-art methods taken form the image foren- sic literature, and on general purpose very deep convolution neural networks (CNN s) properly trained for this task are used.

In image to image translation, images from one domain are mapped to corresponding images in another domain. Using GAN based approach, results have been impressive. In most cases, the network is trained based on a large number of images but in some cases, the correspondences are not known a priori, and no images are available. So an alternate

training procedure is required.

Figure 2: Spot the fake.

Under the adversarial training paradigm, we have two actors: the image-to-image network (generator), and a support network (discriminator). The discriminator tries to tell apart real images from those created by the generator, while the Generator is trained to deceive the discriminator. Input image *A* is first transformed from its original domain*D A* to the output domain*DB* using generator A. Then, image is mapped back to the original domain *A* by generator B. In Fig.2 we show a sample result of this method

applied to satellite images. The image on the left is generated by the algorithm, while the other one is downloaded from Google Maps. At a first glance, no hints of a possible manipulation appear.

Figure 3: Examples of image-to-image translation; Left: Van Gogh or Monet? Right: summer or winter?

In this figure, original image is shown side by side with new versions obtained through a change of context: a different painting style (Van Gogh instead of Monet) or a different season (winter instead ofsummer).We compare a number of methods for the detection of image-to-image translation. Some are methods are:

- GANdiscriminator:
- Steganalysisfeatures:

1) Cozzolino2017:
2) Bayar2016
3) Rahmouni2017:

- DenseNet:

1) InceptionNetv3:
2) XceptionNet:

In order to train and validate the detectors, we need data sets of samples. For each category of detection methods, we use both the real and fake images. So, for example, the *apple2orange* subset includes all the original images of apples and oranges used to train the GAN, and the corresponding fakes (oranges and apples, respectively) generated by the GAN itself once trained.

Table I

SCENARIO 1: RESULTS ON THE ORIGINAL UNCOMPRESSED DATASET.

Accuracy	ap2or	ho2æb	win2sum	citysc.	facades	map2sat	Ukiyoe	Van Gogh	Cezanne	Monet	average
Steganalysis feat.	98.93	98.44	66.23	100.00	97.38	88.09	97.93	99.73	99.83	98.52	94.40
GAN discr.	69.84	90.77	52.31	99.87	98.38	90.44	66.19	96.01	98.97	84.16	83.58
Cozzolino2017	99.90	99.98	61.22	99.92	97.25	99.59	100.00	99.93	100.00	99.16	95.07
Bayar2016	99.26	99.77	50.36	76.02	89.75	79.70	77.23	99.00	98.70	88.89	84.86
Rahmouni2017	88.60	99.20	51.30	72.03	90.38	72.35	97.83	99.97	100.00	97.26	85.71
DenseNet	79.05	95.77	67.68	93.80	99.04	78.30	99.50	97.71	99.93	89.83	89.19
InceptionNet v3	84.95	94.78	58.76	99.41	93.99	70.54	99.80	98.77	99.87	89.89	89.09
XceptionNet	95.91	99.16	76.74	100.00	98.56	76.79	100.00	99.93	100.00	95.10	94.49
average	89.55	97.23	60.58	92.63	95.59	81.97	92.31	98.88	99.66	92.85	89.55

In Table I, for all considered techniques, we show the accuracy obtained for each type of manipulation, together with their average (last column) weighted by the number of sam- ples. Among the considered techniques, the highest average accuracy is obtained by Cozzolino2017. This shallow network provides near-perfect classification for all manipulations ex- cept winter2summer, for an average of 95.07%. However, the handcrafted features also provide very good results, as wellas XceptionNet, among the deep networks. In the last row of **Table I** we also report the average classification accuracy for each manipulation. This allows us to spot the mostchallenging cases, winter2summer and map2sat. On the opposite side, the manipulations most easily detected are horse2zebra, and some painting style transfer.

Table II

SCENARIO 2: RESULTS ON TWITTER-LIKE COMPRESSED IMAGES WITH TRAINING MISMATCH.

Accuracy	ap2or	ho2æb	win2sum	citysc.	facades	map2sat	Ukiyoe	Van Gogh	Cezanne	Monet	average
Steganalysis feat.	50.20	50.12	50.00	50.00	50.00	50.00	50.00	50.20	50.00	49.98	50.05
GAN discr.	50.15	55.24	49.91	50.50	50.38	50.27	48.24	48.90	49.04	47.88	50.19
Cozzolino2017	50.00	50.00	50.00	50.00	50.13	50.05	50.00	50.30	50.00	50.00	50.03
Bayar2016	64.55	81.40	50.05	50.62	84.50	54.15	54.93	89.90	97.37	70.57	67.42
Rahmouni2017	70.43	98.54	50.02	82.71	95.00	50.00	82.43	88.13	100.00	94.42	81.30
DenseNet	76.59	90.32	54.17	65.29	96.63	61.69	96.38	91.56	99.73	87.87	79.76
InceptionNet v3	83.85	86.92	52.67	98.66	93.99	60.00	95.81	93.52	99.57	88.45	85.44
XceptionNet	90.87	96.09	52.56	97.93	98.20	51.34	99.34	97.27	99.70	86.29	87.17
average	67.08	76.08	51.17	68.21	77.35	53.44	72.14	76.22	80.68	71.93	68.92

The good results of the former experiment should be taken with caution. In fact, it is very unlikely that the original fake image generated by a malevolent user is directly available. In **Table II** we show results in the presence of compression forthe very same detectors tested before, that is, the classifiers are still trained on original samples but tested on compressed ones. This simple routine operation impairs dramatically the performance of most detectors under comparison, especially techniques based on handcrafted features or shallow neural networks. Deep networks exhibit a higher robustness, espe- cially XceptionNet, with an accuracy of 87.17%, only 7% worse than the uncompressed case. This indicates that such networks do not rely only on textural micro patterns but also on other features, which survive compression.

Table III

SCENARIO 3: RESULTS ON TWITTER-LIKE COMPRESSED IMAGES.

Accuracy	ap2or	ho2æb	win2sum	citysc.	facades	map2sat	Ukiyoe	Van Gogh	Cezanne	Monet	average
Steganalysis feat.	79.39	90.02	56.66	92.17	73.62	69.39	65.83	95.30	94.73	80.89	81.09
GAN discr.	63.29	91.08	51.90	53.14	88.75	79.35	76.56	80.32	96.41	81.83	73.33
Cozzolino2017	79.57	89.82	53.74	86.81	62.88	89.64	67.67	98.80	99.93	87.33	82.62
Bayar2016	54.64	95.34	50.27	54.00	90.63	52.69	58.90	74.27	99.77	78.60	69.17
Rahmouni2017	84.96	98.35	54.30	57.60	91.88	54.93	96.83	99.63	99.77	89.72	80.97
DenseNet	78.27	93.44	66.94	97.83	98.19	80.45	97.54	98.53	99.57	83.95	88.51
InceptionNet v3	78.60	95.23	64.54	96.09	90.14	63.84	99.53	96.31	100.00	86.21	87.37
XceptionNet	93.52	93.77	67.07	95.11	99.22	67.97	99.66	95.18	99.97	84.02	89.03
average	76.53	93.38	58.18	79.09	86.91	69.78	82.81	92.29	98.77	84.07	81.51

In the last experiment we have considered a worst case, with mismatch between training and test set. When the goal is to detect fakes on social media, it is reasonable to train the classifiers directly on compressed images. The results obtained in this case are reported in **Table III**. The steganalytic features and Cozzolino2017 now perform reasonably well, but more than 10% points below the uncompressed case.

II. CONCLUSION OF GAN:-

From the following experiments, they say that the several detectors for image detection have performed very well on original images and some of them show impairments on compressed images. In the first table, they have shown that the experiments conducted in uncompressed original data have been very impressive. Cozzolino2017 provides near perfect classifications for all except win2sum. The next experiments performed by them is on compressed images. They have used Twitter like compressed images. In this, they performed 2 experiments. In the first experiment, they test the detectors on compressed images with training mismatch. It means that they have trained the detectors on original images but tested on the compressed ones. In this, deep networks have exhibited ahigher robustness, especially XceptionNet, with an accuracyof 87.17%.Inthelastexperiment,againXceptionNethassho wnthe best result. From the above three experiments, they say that test, durability is better preserved by deep networks, especially XceptionNet which keeps working reasonably well even in the presence of training-test mismatching.

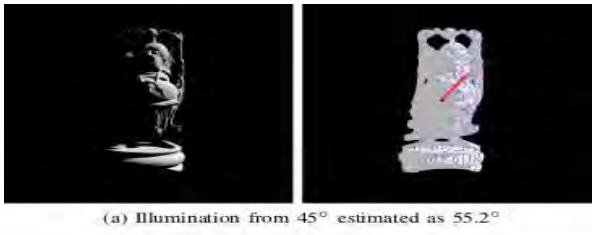
III. GRADIENT:-

Most of the methods for digital-image forensics are sta- tistical. For eg, the fixed pattern noise of a camera can be used to associate an image to a unique source device. Other methods analyze for example JPEG

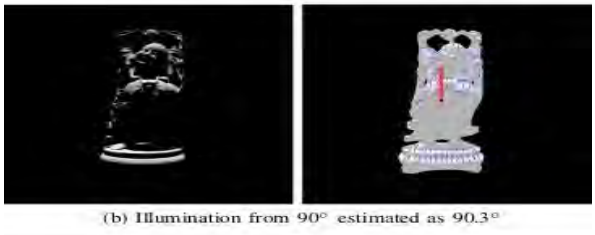
compression artifacts, traces of re sampling, or summarized noise statistics based on steganographic descriptors. Recent works perform similar tasks with

$$d\hat{I}(x,y) = \left(\frac{dI_x(x,y)}{\|dI(x,y)\|}, \frac{dI_y(x,y)}{\|dI(x,y)\|} \right)^{\mathrm{T}}$$

neural networks, e.g., to condition noise patterns on EXIF entries or to directly search for noise inconsistencies. However,it is still an open challenge to reliably apply statisti- cal methods to re compressed and down sampled data Physics- based methods are more resilient to re compression and down sampling. These methods search for physical inconsistencies in images to detect forgeries, like for example in the direction of incident light and the resulting shadows Estimation of the main illumination direction. Image intensity gradients for each masked person are computed and function as input for an estimation of the main light direction. The illumination information can be used as physics-based cue to expose spliced images. The two



(a) Illumination from 45° estimated as 55.2°



(b) Illumination from 90° estimated as 90.3°

(c) Illumination from 135° estimated as 128.6°

persons on the left origin from an image with different illumination as indicated by the estimated main light direction And object shading orthespectral distribution of light Natural objects are much less constrained, but Zheng and Chelappa pointed out that mean gradients still point towards the light source as long as the object is locally spherical.

Tofurthercharacterizethegradientvectorfield,ahistogra m of the gradient vectors is computed. The gradient directions are discretized into 72 bins, which corresponds to an angular resolution of 5 . The bins are collected in a histogram, and the sum of the histogram entries is normalized to 1 to accommodate for different object sizes of A and B. The mean of the normalized gradient vectors, i.e., the dominant lighting direction, is computedas

Where dI(x, y) denotes the magnitude-normalized gradient vectors, computed as

The normalization leads to the effect that only thedirection ofthevectoristakenintoaccount[31].Itcontributestoahi gh robustness to changes in albedo, and mitigates the need for intrinsic image decomposition for albedo neutralization. The mean in Eqn. 14 is used as the estimate for the dominant illuminant direction in the image plane they compute the dominant lighting directions a and b for two objects Aand

B. To quantify differences in illumination, they use the cosine dissimilarity LD,

$$L_D(\boldsymbol{a},\boldsymbol{b}) = 1 - \frac{(c(\boldsymbol{a},\boldsymbol{b}) + 1.0)}{2} \quad ,$$

where

$$c(\boldsymbol{a},\boldsymbol{b}) = \frac{\boldsymbol{a}^{\mathrm{T}} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|}$$

is the cosine of the angle between the two illuminant directions. This measure ranges from 0 to 1, with value 0for identical directions, and 1 for opposite directions. The proposed features, namely the differences in tilt LD, the divergenceoftiledestimatesDD,individualdifferences ofthe tiled estimates TD, and the correlation between the gradient histograms HD are used to quantify the dissimilarity of two lighting environment. This method being more robust than previous lighting based methods can be fully automated.They use the popular Mask R-CNN neural network to segment the image, which provides a coarse pixel-wise segmentation of object instances. Objects with an area below 1% of the image pixels are excluded from the analysis, since these consist of too fewpixels.

For splicing detection, they make the common assumptionthat objects from different images exhibit different lighting environments. Thus, the classification task is to determine whether two objects stem from the same image or from two different images.

Block size (pixels)	DA (%)	FPR_1 (%)	FPR_2 (%)
5 × 5	98.5186	6.2592	2.7709
10 × 10	98.4263	6.1293	2.3956
15 × 15	98.4092	6.1129	2.4251
20 × 20	98.4007	6.0382	2.8134
25 × 25	98.3770	6.0647	2.4074
30 × 30	98.3089	6.0184	2.4692
35 × 35	98.2972	5.9374	2.6008

Results of Gradient Based processing:

Dominant lighting conditions:
Themedianerrorforalllightdirectionsis20.Theothermethods performslightlyworsewithmedianerrorsforalllightdirections between 22 and 26 .

Fully automated splicing detection:

1) DSO-1 Dataset: Statistical methods perform betterthan the physics based methods because dataset contains images captured with frontal flash. However, after resizing and com- pressing the dataset, the performance of statistical methods drops significantly and their method performsbest.

2) OIS:The source images of thedataset

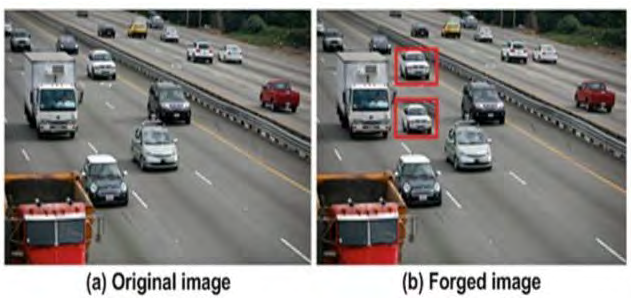
Stem from the publicly available OpenImages V4 dataset, which contains about nine million images annotated with image-level labels and bounding-boxes.only consider thesplic- ing of persons for manipulationdetection.

Inthiswork,theauthorsgaveahighlyresilientphysicsbased method that can be fully automated for comparing 2d lighting environments, which unlike previous lighting based methods can be applied to wider range of images. The method is based on four features computed on gradients of an object. The descriptor uses the whole object area instead of a single noisy area which contributes to its higher performance than previous methods.A simple logistic regression on the proposed features can be deployed to distinguish between objects from identical or different images. Fully automated segmentation and classification achieves an AUC of up to 0.774 onsplicedimages.

One of the most remarkable quality of this descriptor is that it remains unaffected by compression and resizing where statistics based methods fail and better than previous lighting based methods.

IV. COPY-MOVE FORGERY DETECTION EXPLOITING STATISTICAL IMAGEFEATURES:-

Many times attackers apply some post-processingoperations (such as blurring, edge smoothing, noise addition and filtering of the forged region, etc.) on the forged image, so that the manipulated regions seem natural.Here, we present a region duplication detection method for images, which utilizes statistical image features in thefrequency domain, viz., mean and variance [3]. The primary motivation is to exploit the local statistics of an image, using block-based mean and variance computations. The contribu- tion of each individual block with respect to the entire image, in terms of intensity, can be measured using the image mean, whereas the variation of individual pixel intensities within an image sub block, is indicated by the variance



V. CONCLUSION OF COPY-MOVE FORGERY DETECTION:-

Inthischapter,wepresentaregionduplicationforgerydetection methodutilizingmeanandvariancestatisticalfeatures.Thoseare simple yet powerful statistical image features and effective for forgery detection also. The presented

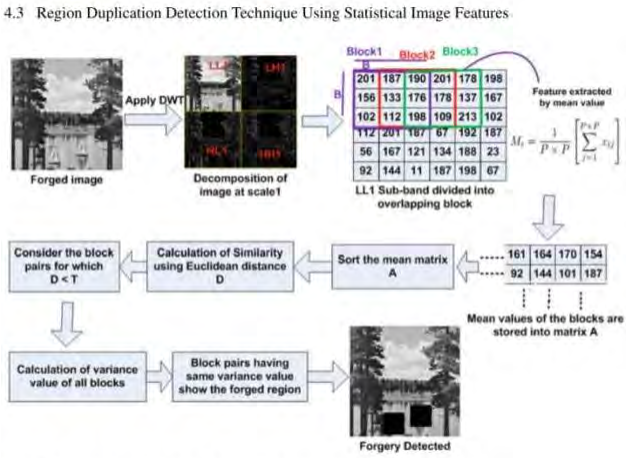
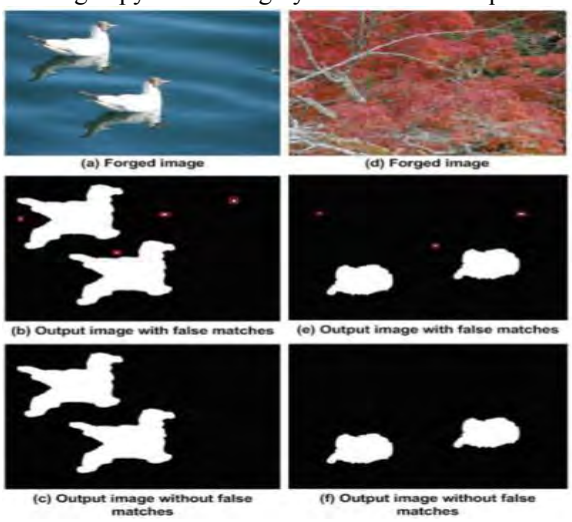


Fig. 4.2 Operational steps of the proposed algorithm

method operates by dividinganimageintopixelsub-blocksinthefrequencydomain, and also deals with the statistical features of all image blocks. The experimental results verify the efficacy of the method and show the supremacy of the technique over the existing copy- move forgery detection techniques.



VI. CONCLUSION:-

Wehavepresentedareviewoftheresearchcarriedoutinthearea of detection of fake or tampered images. Furthermore, a number of algorithms were discussed

which can be used to detect those images.Manipulatedimagesaremainlyfoundintwoforms: one intheiroriginalformandsecondintheircompressedformwh ich are shared on social media platforms like Twitter. We have seen three different algorithms used for detection. In this first one, deep network methods like XceptionNet has shown the best results throughout the three cases. Its performance is consistent across original manipulated image and on compressed manipulated image and also fares better when tested on compressed data but trained on original data. The gradient method uses a physics based approach. It can be fullyautomated for comparing 2D lighting environment which can be applied to a wide area unlike other lighting methods. Unlike detection methods for GAN which shows slightly different results for original and compressed images, gradient methods remain unaffected by compression and resizing. The last method is a statistical method which detects copy move forgery. The region duplication forgery detection algorithm using mathematical methodssuchasmeanandvariancewasused.Afteranalysis, the accuracyashighas98.51%wasrecorded.Theaccuracywas very high for the small blocks of pixels which went on to reduce marginallyasthesizeoftheblockincreased.Thesametrend was observed in the false positive rate which was recorded at 6.25% at the highest. Thus, it can be concluded that image forgery can be detected with high accuracy for small to medium block size image using the basics ofstatistics.

VII. ACKNOWLEDGMENT

1. Francesco Marra, Diego , , Luisa , Detection of GAN-generated Fake Images over Social Networks,2018.

2. EstebanAlejandroArmasVega,Edgar,AnaLucilaSandovalOrozco, Luis , Passive Image Forgery Detection Based on the Demosaicing Algorithm and JPEG Compression (ongoing research).

3. Roy, , , , Copy-Move Forgery Detection ExploitingStatistical Image Features, May2019.

Predicting Customer Purchase Behavior on Black Friday

Devesh Varma , Angela Roy , Harsh Ramani , Abhishek Yadav
Third Year, Department of Computer Engineering
Thakur College of Engineering and Technology,
Mumbai, India

Abstract- **It is no surprise that businesses spend millions of dollars in carrying out market research before coming up with a new service or product. Despite that, it is important to recognize that the final product doesn't sell itself and actually requires the right marketing tools or personalized marketing to make itself visible to the potential consumers. With millions of buying options at our disposal, the behavior of this era's consumers keeps flickering, and it is important to advertise the right prod-uct to the right customer at the proper time.**

The project is built with to accurately predict how much will a user spend on a Black Friday sale given his de-tails. The data for our model is a sample of the transac-tions made in a retail store. We intend to know better the customer purchase behaviour against different products. Specifically, here the problem is a regression problem where we are trying to predict the dependent variable (the amount of purchase) with the help of the infor-mation contained in the other variables.

The features of customer like age, marital status, occu-pation, Current city, will help us accurately predict the amount he/she will spend on a Black Friday Sale. Root mean square Error (RMSE) is used for performance measure. RMSE is commonly used general purpose er-ror metric.

Key Terms- Black Friday, Shopping, Prediction, Multivariate Linear Regression

I. INTRODUCTION

A. Background

The term Black Friday is known to shoppers and retail-ers in the USA as both the day after the Thanksgiving holiday and one of the busiest shopping days of the year the day is characterized by high demand for retail goods. This makes stores to reduce the price on most of its products in order to attract customers due to increased competition. As a result, many stores offer highly pro-moted sales on Black Friday and open very early, such as at midnight, or may even start their sales at some time on Thanksgiving. Black Friday is not an official holiday, but some states observe "The Day after Thanksgiving" as a holiday for state government employees, some-times in lieu of another federal holiday, such as Colum-bus Day [1].

Black Friday has routinely been the busiest shopping day of the year in the United States since more than a decade [2].
With the use of recent development in Data Science and Machine Learning we can use the details of a person to find his purchase habits, predict his future purchases and use this to give personalized offers to them. This was not possible a few decades ago where the technology was not as advanced as it was today.

B. Problem Definition

Given a data of users such as Gender, Age, City loca-tion, Marital Status, Money spent and products bought, build a model to predict the amount of money he will be spending on a Black Friday sale. The main goal is to accurately predict the amount a person spends and the secondary goal is to predict the major factors influenc-ing his spending habits from the given data.

II. LITERATURE SURVEY

A. Machine Learning for Prediction

Machine learning is a subset of artificial intelligence (AI) which learns from experience with respect to some class of tasks, thereby improving its performace with experience. that provides system the ability to automat-ically learn and improve from experience without hav-ing someone programming it. The main goal of machine learning is to focus on the development of computer programs that can access data and use it learn for them-selves. The process of learning begins with observa-tions, facts, information or data, such as examples, di-rect experience, or instruction, in order to look for pat-terns in data and make better decisions in the future based on the examples that are given by the user.

The main goal is to allow the computers learn automati-cally without human intervention, help or assistance and adjust actions accordingly. Machine learning with the help of suitable algorithm enables analysis of large quantities of data. Although it generally gives faster and accurate results, it also requires addition resources such as CPU time and memory time to do the necessary pro-cessing and get accurate results [3].

Predictive analytics is a part of Machine Learning which is mainly driven by predictive modelling. It's more of an approach than a process. Predictive analytics and ma-chine learning go hand-in-hand; as predictive models typically include a machine learning algorithm. These models can be trained over time to respond to new data or values, delivering the results the business needs. Pre-dictive modelling largely overlaps with the field of ma-chine learning.

There are two types of predictive models. They are Classification models, that predict class membership, and Regression models that predict a number. These models are then made up of algorithms. The algorithms perform the statistical analysis, determining trends and patterns in data and predicting the values based upon them. Predictive analytics software solutions have built in algorithms that is used to make predictive models which help in predicting the most accurate value. The algorithms are defined as 'classifiers', identifying which set of categories data belongs to [4].

B. Multivariate Linear Regression Model

The goal of data analysis is to extract meaningful data or information from raw data with accurate estimation. One of the most important and common question con-cerning if there is statistical relationship between a re-sponse variable (Y) and explanatory variables (Xi). An option to answer this question is to employ regression analysis in order to *model* its relationship. Further it can be used to *predict* the response variable for any arbitrary set of explanatory variables.

Multivariate Regression is one of the simplest Machine Learning Algorithm. It comes under the class of Super-vised Learning Algorithms i.e, when we are provided with training dataset [5].

Multiple/Multivariate linear regression models the rela-tionship between two or more known variables and a re-sponse variable by fitting a linear equation to observed data. Every value of the independent variable/varia-bles x is associated with a value of the dependent varia-ble y.

III .ARCHITECTURE

	User_ID	Product_ID	Gender	Age	Occupation	City_Category
2	1000001	P00069042	F	0-17	10	A
3	1000001	P00248942	F	0-17	10	A
4	1000001	P00087842	F	0-17	10	A
5	1000001	P00085442	F	0-17	10	A
6	1000002	P00285442	M	55+	16	C
7	1000003	P00193542	M	26-35	15	A
8	1000004	P00184942	M	46-50	7	B
9	1000004	P00346142	M	46-50	7	B
10	1000004	P0097242	M	46-50	7	B
11	1000005	P00274942	M	26-35	20	A

Stay_In_C	Marital_St	Product_C	Product_C	Product_C	Purchase
2	0	3			8370
2	0	1	6	14	15200
2	0	12			1422
2	0	12	14		1057
4+	0	8			7969
3	0	1	2		15227
2	1	1	8	17	19215
2	1	1	15		15854
2	1	1	16		15686
1	1	8			7871

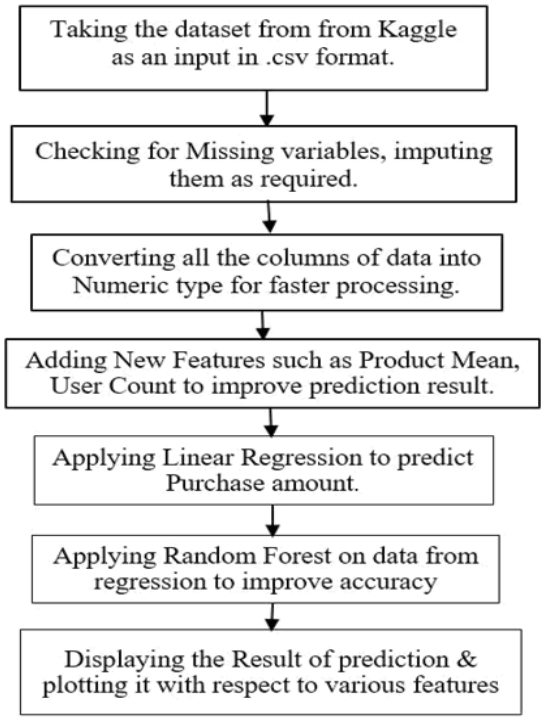
Fig. 1 Structure of Input Dataset

The analysis of the Black Friday data set used in this paper includes 12 variables: Product_ID, Gen-der, Age, City_Category, Stay_In_Cur-rent_City_Years, Marital_Status, Product_Cate-gory_1 and Purchase.

This data set is cleaned from the original data set which contained 3 additional variables - Product_Cate-gory_2 and Product_Category_3. Product_Cate-gory_2 and Product_Category_3 contain missing obser-vations which are imputed as zero.

The Black Friday data set is further cleaned by stating the format of each variable. This included chang-ing Product_ID, Gender, Age, City_Category, Mari-tal_Status and Product_Category from character varia-bles to integers. These variables are nominal since they do not have order or a distance metric. 'Age' is ordinal because the age groups are increasing with actual age, but are not equally spaced.

IV. FLOW OF THE PROJECT



V. IMPLEMENTATION

Source Code:

```
library(caTools)
library(plyr)
library(data.table)
library(DataExplorer)
library(dplyr)
library(ggplot2)
library(vcd)
library(rpart)
library(randomForest)
#Reading Data from CSV file

blackFriday <-
read.csv("H:/train1.csv",stringsAsFactors = F)
blackf_data <- fread("H:/train1.csv")

# STEP 1: DATA
OVERVIEW
head(blackFriday,10)
#What is the structure of the
data? str(blackf_data) # Text
#plot_str(blackf_data) #
Network Graph
#STEP 2: DATA
PREPROCESSING #Filling
missing data with zeroes
blackFriday[is.na(blackFrid
ay)]<-0

#Converting all variables to Numeric type
blackFriday$City_Category[blackFriday$C
ity_Cate-gory=="A"]<-"1"
                                                    3
blackFriday$City_Category[blackFriday$City_Cate
-
gory=="B"]<-"2"
blackFriday$City_Category[blackFriday$City_Cate
-
gory=="C"]<-"3"
blackFriday$City_Category<-as.integer(blackFri-
day$City_Category)

# converting age variable to numeric
blackFriday$Age[blackFriday$Age == "0-17"]
<- "15" blackFriday$Age[blackFriday$Age ==
"18-25"] <- "21"
blackFriday$Age[blackFriday$Age == "26-
35"] <- "30"
blackFriday$Age[blackFriday$Age == "36-
45"] <- "40"
blackFriday$Age[blackFriday$Age == "46-
50"] <- "48"
blackFriday$Age[blackFriday$Age == "51-
55"] <- "53"
blackFriday$Age[blackFriday$Age == "55+"]
<- "60" blackFriday$Age <-
as.integer(blackFriday$Age)

# converting stay in current city to numeric
blackFriday$Stay_In_Current_City_Years[
blackFri-day$Stay_In_Current_City_Years
== "4+"] <- "4"
blackFriday$Stay_In_Current_City_Years
<-as.inte-
ger(blackFriday$Stay_In_Current_City_Y
ears)

# converting gender to binary
blackFriday$Gender <- ifelse(blackFriday$Gender
== "F", 1,
0)
str(blackFriday)
blackf_data<-blackFriday
#STEP 3: FEATURE
ENGINEERING #Adding
Product_Mean Feature to our data
product_mean <- ddply(blackFriday, .(Product_ID),
summa-rize, Product_Mean=mean(Purchase))
```

```
blackFriday <- merge(blackFriday, product_mean,
by="Prod-uct_ID")

# feature representing the count of each user
user_count <- ddply(blackFriday,
.(User_ID), nrow) names(user_count)[2]
<- "User_Count"
blackFriday <- merge(blackFriday, user_count,
by="User_ID")

#feature representing the count of each product
product_count <- ddply(blackFriday, .(Product_ID),
nrow)
head(product_count)
names(product_count)[2] <- "Product_Count"
blackFriday <- merge(blackFriday, product_count,
by="Prod-uct_ID")

#feature representing the proportion of times the
user pur-chases the product more than the product's
average blackFriday$flag_high <-
ifelse(blackFriday$Purchase >
blackFriday$Product_Mean,1,0)
user_high <- ddply(blackFriday, .(User_ID),
summarize, User_High=mean(flag_high))
blackFriday <- merge(blackFriday, user_high,
by="User_ID")

dt = sort(sample(nrow(blackFriday),
nrow(blackFriday)*.8))
train<-blackFriday[dt,]
test<-blackFriday[-dt,]

model<-
      lm(Purchase~Gender+Stay_In_Current_City
      _Years
      +Product_Category_2+Product_Category_3
      +Product_Mean+User_Count+Prod-
uct_Count+flag_high+U
      ser_High
      +Occupation,data
      =train)
summary(model)
#prediction<-predict(model,test)
#blackFriday      <-      merge(blackFriday,
                  prediction,
by="User_ID")
model<-
randomForest(Purchase~Gender+Age+Marital_Sta-
tus+Stay_In_Current_City_Years
            +Product_Category_1+Prod
uct_Cate-gory_2+Product_Category_3
            +Product_Mean+User_Cou
nt+Prod-
uct_Count+flag_high+User_High
            ,data=train)
plot(model)
pred<-predict(model,test)
class(test)
class(train)
View(test)
View(train)
print(class(pred))
#Tweaking Final Prediction so that the prediction
remains in
the range
pred[pred<187]<-187
pred[pred>23949]<-23949
final<-cbind(Actual=test$Purchase,Predicted=pred)
as.data.frame(final)->final
View(final)
max(test$Purchase)
min(test$Purchase)
(test$Purchase-final$Predicted)->error
cbind(test,error)->test
rmse<-sqrt(mean(test$error^2))
summary(model)
rmse

blackf_data$User_ID <-
as.factor(blackf_data$User_ID)
blackf_data$Product_ID <-
as.factor(blackf_data$Prod-uct_ID)
blackf_data$Gender                    <-
as.factor(if_else(blackf_data$Gender == 'M', 'Male',
'Female'))
blackf_data$Age <- as.factor(blackf_data$Age)
blackf_data$Occupation                <-
as.factor(blackf_data$Occupa-tion)
blackf_data$City_Category<-           as.fac-
tor(blackf_data$City_Category)
blackf_data$Stay_In_Current_City_Years <- as.fac-
tor(blackf_data$Stay_In_Current_City_Years)
blackf_data$Marital_Status      <-    as.fac-
tor(if_else(blackf_data$Marital_Status    ==    1,
'Married', 'Sin-gle'))
# Due to the large dataset, the average values were
used for this analysis (using pipe operator)
new_data <- blackf_data %>%
  group_by(User_ID, Age, Gender, Occupation,
City_Cate-gory, Stay_In_Current_City_Years,
Marital_Status) %>%



  summarise_each(funs(mean),
Product_Category_1, Prod-uct_Category_2,
Product_Category_3, Purchase)

# Rename the average values accordingly
colnames(new_data)[8] <- "Product_Cat_1_Avg"
colnames(new_data)[9] <- "Product_Cat_2_Avg"
```

```
colnames(new_data)[10] <- "Product_Cat_3_Avg"
colnames(new_data)[11] <-
"Avg_Purchase_Amount"
ggplot(new_data, aes(Age,
Avg_Purchase_Amount, fill = Gender)) +
geom_col() + facet_wrap(~ Gender) +
  labs(title = "Age Group/Gender Vs Average
Purchase Amount")

#   City Category versus Average purchase amount
ggplot(new_data, aes(City_Category, Avg_Pur-
chase_Amount, fill = City_Category)) +
geom_col() +
labs(title = "City Category Vs Average Purchase
Amount")

#   Stay in current city versus Average purchase
amount
ggplot(new_data,
aes(Stay_In_Current_City_Years,
Avg_Purchase_Amount,          fill          =
Stay_In_Current_City_Years)) + geom_col() +
labs(title = "Stay in current city Vs Avg_Pur-
chase_Amount")
#   Marital status versus Average purchase amount
ggplot(new_data, aes(Marital_Status, Avg_Pur-
chase_Amount, fill = Marital_Status)) +
geom_col() +
labs(title = "Marital Status Vs
Avg_Purchase_Amount")
```

VI. RESULTS AND DISCUSSION

	Actual	Predicted
7	6187	5959.1365
16	15854	15854.9660
32	5293	4370.1015
34	6914	8411.8718
38	7381	7683.1275
47	19614	14659.7044
48	9872	10107.2901
49	19203	18826.6010
54	11877	14204.1258
56	16662	16446.5617
60	10946	12680.8579
64	12642	15482.4698
66	19681	16202.0383
67	8718	8614.0623

Fig. 2 Actual value v/s Predicted values

The Root Mean Squared Error (RMSE) is:

```
# Age group versus Average purchase amount
```

```
> rmse
[1] 1711.175
```

Gender played a major role in the Amount of Money spent. Men spent three times more when compared to Women

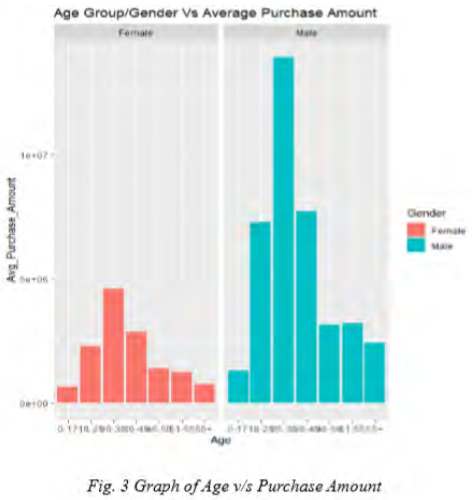


Fig. 3 Graph of Age v/s Purchase Amount

Fig. 3 Graph of Age v/s Purchase Amount

Comparion of which city spent the most money:

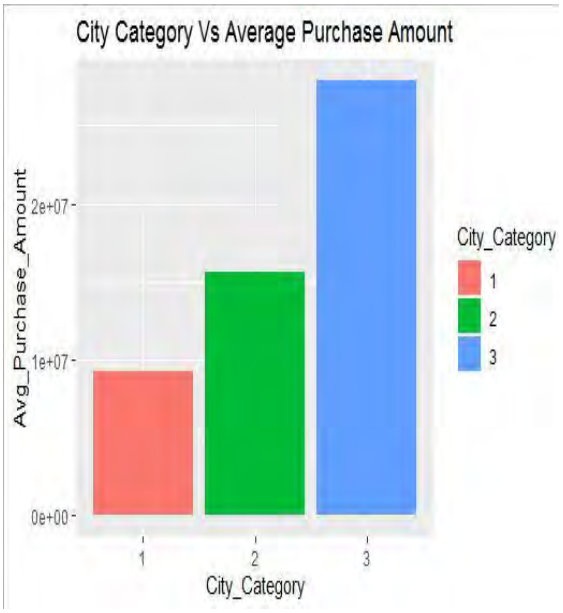


Fig. 4 Graph of City Category v/s Average Purchase Amount

Fig. 5 Relation between Marital Status and Purchase Amount

Thus from our results we have proved some features of a person demographic to decide on how much he spends whereas some features had neglible outcome on the Pur-chase amount.

References

[1] "Retailers & Sales: Using Social Listening to Learn More About Black Friday – Brandwatch". Brandwatch. December 2, 2014.

[2] International Council of Shopping Centers. "Holi-day Watch: Media Guide 2006 Holiday Facts and Fig-ure". Archived from the original (PDF) on November 29, 2010. ShopperTrak, Press Release,"ShopperTrak Reports Positive Response to Early Holiday Promotions Boosts Projections for 2010 Holiday Season". Archived from the original on November 29, 2010. Re-trieved 2012-01-23. (November 16, 2010).

[3] Marco Varone, Daniel Mayer "What is Machine Learning? A definition" http://www.expert-system.com/machine-learning-definition

[4] Katrina Wakefield "Predictive analysis and machine learning" http://www.sas.com/en_gb/insights/arti-cles/analytics/a-guide-to-predictive-analytics-and-ma-chine-learning.html

[5] Abhishek Sharma "Multivariate Regresion" http://www.geeksforgeeks.org/multivariate-regression/

Artificial Intelligence

Rajvi Shah , Harsh Pathak , Pradeep Pal
Thakur College of Engineering and Technology
Maharashtra, India

Abstract**: Artificial Intelligence (A.I.) may be a multidisciplinary field whose goal is to automate activities that presently require human intelligence. Recent successes during AI. It includes computerized medical diagnosticians and systems that automatically customize hardware to particular user requirements. the main problem areas addressed during AI. It is often summarized as Perception, Manipulation, Reasoning, Communication, and Learning. Perception cares with building models of the physical world from sensory input (visual, audio, etc.). Manipulation cares with articulating appendages (e.g., mechanical arms, locomotion devices) so as to affect a desired state within the physical world. Reasoning cares with higher level cognitive functions like planning, drawing inferential conclusions from a world model, diagnosing, designing, etc. Communication treats the matter understanding and conveying information through the utilization of language. Finally, Learning treats the matter of automatically improving system performance over time supported the system's experience. Many important technical concepts have arisen from A.I. that unify these diverse problem areas which form the inspiration of the science. The system must automatically organize and utilize this information to unravel the precise problems that it encounters. This organization process is often generally characterized as an enquiry directed toward specific goals. The search is formed complex due to the necessity to work out the relevance of data and since of the frequent occurrence of uncertain and ambiguous data. Heuristics provide the A.I. system with a mechanism for focusing its attention and controlling its searching processes.**

Keywords: human intelligence, sensory input, arisen, articulating appendages.

I. INTRODUCTION

Artificial Intelligence is an approach to form a computer, a robot, or a product to think how smart human think. AI may be a study of how human brain think, learn, decide and work, when it tries to unravel problems. and eventually this study outputs intelligent software systems. The aim of AI is to enhance computer functions which are associated with human knowledge, for instance , reasoning, learning, and problem-solving.

The objectives of AI research are reasoning, knowledge representation, planning, learning, tongue processing, realization, and skill to maneuver and manipulate objects. There are long-term goals within the general intelligence sector. the normal problems (or goals) of AI research include reasoning, knowledge, planning, learning, tongue processing, perception, explain ability and therefore the ability to maneuver and manipulate objects. General intelligence is among the field's long-term goals.

Approaches include statistical methods, computational intelligence, and traditional symbolic AI. Many tools are utilized in AI, including versions of search and mathematical optimization, neural network sand methods supported statistics, probability and economics. The AI field draws upon computing , mathematics, psychology, linguistics, philosophy, neuroscience, artificial psychology and lots of others.

Approaches include statistical methods, computational intelligence, and traditional coding AI. During the AI research associated with search and mathematical optimization, artificial neural networks and methods supported statistics, probability, and economics, we use many tools. computing attracts AI within the field of science, mathematics, psychology, linguistics, philosophy then on.

Generally, A.I. systems function supported a knowledge domain of facts and rules that characterize the system's domain of proficiency. the weather of a knowledge domain contains independently valid (or a minimum of plausible) chunks of data . The necessarily adaptive organization of A.I. systems yields the need for A.I. computational Architectures. All knowledge utilized by the system must be represented within such an architecture. The acquisition and encoding of real-world knowledge into A.I. architecture comprises the subfield of data Engineering.

Use of machine intelligence to capture user interests and online behavior is very common now on the web. Whenever user visits a web site, in the background some intelligent algorithms execute to capture data about user’s interests and online behavior. The browsing history is analyzed by those systems and information is extracted from the data to make an intelligent guess to display advertisements of products that are of interest to the user.

II. ARTIFICIAL INTELLIGENCE RESEARCH AREAS

The working domain of AI is large in breadth and breadth. Therefore before proceeding further considers the prospering and customary research areas within the domain of AI are:-

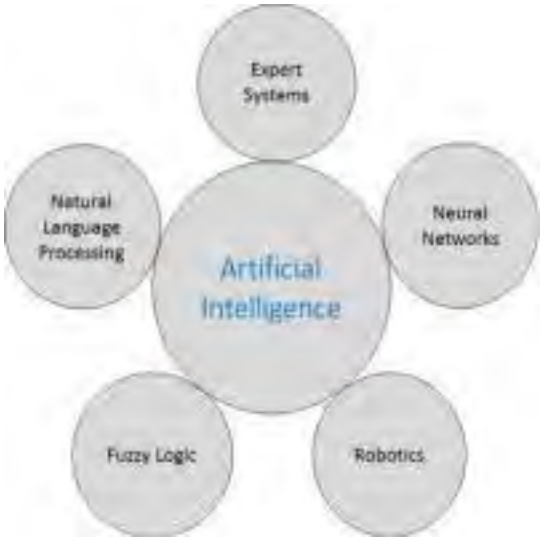


Figure 1: AI Research area

Expert System - In AI , an expert system are used for solving complex problems by reasoning about knowledge, represented primarily by if-then rules instead of by conventional procedural code. In general, an expert system may be a computing system that uses the decision-making capability of a person's expert.

Neural Networks - Neural networks are system of interconnected ? neurons? which exchange messages between each other. In machine learning artificial neural networks (ANNs) belongs to a family of model inspired by biological neural networks (the systema nervous of animals, present inside a brain) and are used for approximate functions or estimate a large number of inputs which are generally unknown.

Robotics - Robotics is a branch of Artificial Intelligence (AI), it is mainly composed of electrical engineering, mechanical engineering and computer science engineering for construction, designing and application of robots. Robotics is science of building or designing an application of robots. The aim of robotics is to style an efficient robot.

Fuzzy logic - symbolic logic was introduced in 1965 as a proposal of fuzzy pure mathematics . It is applied to varied fields, from AI to regulate theory. Fuzzy logic could also be a sort of many-valued logic during which truth table values of variable may be real between 0 and 1.

Natural Language Processing - Natural language processing (NLP) is a method of communicating with an intelligent system by using a natural language such as English. The input and output of NLP system is speech and transcription .

III. TYPES OF ARTIFICIAL INTELLIGENCE

A. Purely Reactive:

The most basic sorts of AI systems are purely reactive, and have the power neither to make memories nor to use past experiences to tell current decisions. Deep Blue, IBM's chess-playing supercomputer, which beat International Grandmaster Garry Kasparov within the late 1990s, is that the perfect example of this sort of machine.

Deep Blue can identify the pieces on a chess board and skills each moves. It can make predictions about what moves could be next for it and its opponent. And it can choose the foremost optimal moves from among the chances .

B. Limited memory:

machines can look into the past. Self-driving cars do some of this already. For example, they observe other cars' speed and direction. That cannot be wiped out a only one moment, but rather requires identifying specific objects and monitoring them over time. These observations are added to the self-driving cars' preprogrammed representations of the planet , which also include lane markings, traffic lights and other important elements, like curves within the road. They're included when the car decides when to vary lanes, to avoid isolating another driver or being hit by a close-by car.

C. Theory of mind:

We might stop here, and call now the important divide between the machines we've and therefore the machines we'll integrate the longer term . However, it's better to be more specific to debate the kinds of representations machines got to form, and what they have to be about.

Machines within the next, more advanced, class not only form representations about the planet , but also about other agents or entities within the world. In psychology, this is often called "theory of mind" – the understanding that folks , creatures and objects within the world can have thoughts and emotions that affect their own behavior.

IV. FEATURES OF AI

Artificial intelligence is so revolutionary, that some people consider it because the new age electricity. When Franklin first discovered electricity, it had been thought of as extremely profound- something that changed our lives forever. Scientists and researchers are regarding AI to be equally revolutionary-something which will transform the way most industries work.

- Eliminate dull and boring tasks
- Data ingestion
- Imitates human cognition
- Futuristic
- Prevent natural disasters
- Facial Recognition and Chatbots

V. PROS OF ARTIFICIAL INTELLIGENCE

- Minimal Human Intervention
- Faster and Accurate

VI. CONS OF ARTIFICIAL INTELLIGENCE

- This can prevent sympathizing with emotions for human contact, such as in being nurses. This can also reduce wisdom can understanding.
- This can prevent common sense occuring. Even if system is coded with common sense,it will be difficult for the sytem to get the common sense as that human can.
- Robots, with them replacing jobs, unemployment will increase.
- As seen partially with smartphones and other technology already, humans can become too dependent on AI and lose their mental capacities.
- If someone try to harm anyone using system that will lead to the destruction.
- AI as robots can take place of humans and enslaving us.

VII. FUTURE SCOPE

Artificial intelligent is a system which act like human. Basically it is not human but it will work like us..think like us react like us…

In future there will be time come when human don't need to work. System will teach the student with experimental things and in intelligent way so that student can understand the things properly.

Artificial Intelligence is to make replicate of human brain's capabilities so that the system start doing all those activities that the human is doing and in much less time. Trends of AI usage are changing day by day. Across the world, many companies are investing their energy, money and time in developing AI systems. However, it must be noted that business persons are not the only investors in AI technology. According to a report by The New Yorker, the government of China has been quite aggressive in using AI technology to control innovation in the near future.

VIII. CONCLUSION

Artificial Intelligence and therefore the technology are one side of the life that always interest and surprise us with the new ideas, topics, innovations, products etc. AI remains not implemented because the films representing it(i.e. intelligent robots), however there are many important tries to succeed in the extent and to compete in market, like sometimes the robots that they show in TV. Nevertheless, the hidden projects and therefore the development in industrial companies. At the end, we've been during this research through the AI definitions, brief history, applications of AI publicly , applications of AI in military, ethics of AI, and therefore the three rules of robotics. This is not the top of AI, there's more to return from it, who knows what the AI can do for us within the future, maybe it'll be a whole society of robots.

References

[1] Robot rightsTony Colombo on Medium link https://medium.com/pensieri-in- liberta/robot-rights- 7bbd03baacfd:

[2] Artificial Intelligence and Robot rights by kurzgesagt.

[3] Artificial Intelligence by School of life. Link for the video: https://www.youtube.com/watch?v=9TRv0cXUVQw

[4] Artificial Intelligence article by cold fusion

Interpretation of Models on the Onset of Diabetes using Machine Learning

Anshika Pandey , Sakshi Singh , Anand Khandare

Thakur College of Engineering & Technology, Mumbai.

anshikapandey603@gmail.com,sakshisaku3000@gmail.com , anandkhandare@thakureducation.org

Abstract-- **Diabetes mellitus is one of the most serioushealthchallengesfaced bytheworldtoday. Theobjectiveistodiagnosticallypredictwhetheror not a patient has diabetes, based on certain diagnostic parameters included in the dataset. Under the supervised learning, we have used Logistic Regression and Decision Tree. Using Decision Tree, the accuracy came out to be approximately 69%. In order to further test the accuracy, Logistic Regression predicted the accuracy to be 78%. The prediction is furtherdone and graphically represented through K-Meansand R-SquaredMethod.Thetesterrorandaccuracyfor logistic regression are then calculated using ConfusionMatrix.Theperformanceofclassifieris evaluated based on precision, recall, accuracy and is estimated over correct and incorrectinstances.**

Keywords-- Diabetes Mellitus, Machine Learning, Logistic Regression, Decision Tree, K-Means.

I. INTRODUCTION

Diabetes mellitus is a kind of metabolic diseases in which a person has high blood sugar. There are two general reasons for diabetes: 1- the body does not produce enough insulin. Only 5-10% of people with diabetes have this form of the disease (type1). With the help of insulin therapy and other treatments, even young children with type1 diabetescanlearntomanagetheirconditionandlive healthy. 2- Cells do not respond to the insulin that is produced (type2). Insulin is the most essential hormone that regulates uptake of glucose from the blood into most cells (not central nervous system cells).Thus,deficiencyofinsulinplaysacentralrole in all forms of diabetes mellitus. Type 2 diabetes mellitus is a complex disease of majorhealthImportance. This incidence is rapidly increasing in the developed countries. It is estimated by the year 2030, there will be 366 million people affected by Type 2 diabetes worldwide. Medical diagnostics is quite difficult task which is mostly done by expert doctors.Anexpertdoctorcommonlytakesdecisions by evaluating the current test results of a patient or compares the patient's results with other patients with the same condition by referring to the previous decisions.Thus, itisverydifficultforaphysicianto diagnose hepatitis. For this reason, in recent times, many machine learning and data mining techniques have been considered to design automatic diagnosis system fordiabetes.

Machine learning is an upcoming new technology with a wide range of applications. It has the potential to become one of the key components of intelligent information systems, inferred from large databases of recorded information that is to be applied as knowledge in various practical ways—such as being embedded in automatic processes, or useddirectlyforcommunicatingwithhumanexperts and for educational purposes. Most of the research work is directed towards the invention of new algorithms for learning, rather than towards gaining experience in applying existing techniques to real problems.

II. RELATEDWORK

Design of the prediction models for diabetes diagnosis has been an area of active research for the past decade. Most of the models found are based on clustering algorithms. In the proposed project the employed three techniques namely are: Regression algorithm, K-means clustering and Classification algorithm, for the classification of the patients suffering from diabetes mellitus. The performance

ForK-meansprovedtobebetterthanotherswhenall the similar symptoms were grouped into clusters using these algorithms. In recent times, a study conducted is intended to discover the hidden knowledge from a particular dataset to improve the quality of healthcare for the diabetic patients. This paper approaches to aim the diagnoses of diabetes by using Machine Learning and demonstrated the need for pre-processing and replacing missing values in the dataset. Through the learned training set, a better accuracy was achieved with lessertime. Hence, there is a requirement of a model that canbe developed that provides reliable, faster and cost effective methods to provide information related to the probability of patients suffering from diabetes. In the present work, an attempt is made to analyses the

diabetes parameters and to establish a probabilistic relation between them using Machine LearningModels.

III. PROPOSEDMETHODOLOGY

A. LogisticRegression

Logistic regression is an applied mathematics model that in its basic type uses a logistical operate to model a binary variable. In multivariate analysis, logistical regression (or logit regression) is estimating the parameters of a logistical model (a type of binary regression). Mathematically, abinary logisticalmodelincludesavariablewith2attainable values, like pass/fail that is portrayed by associate indicator variable, wherever the 2 values area unit labelled"0"and"1".Withinthelogisticalmodel,the log-odds (the exponent of the odds) for the worth labelled "1" may be a linear combination of 1 or a lot of freelance variables ("predictors"); the freelance variables will everbe a binary variable (two categories, coded by associate indicator variable) or a nonstop variable (any real value). The corresponding likelihood of the worth labelled "1" will vary between zero (certainly the worth "0")and one(certainlytheworth"1"),hencethelabelling;the operate that converts log-odds to likelihood is that the logistical operate, thence thename.

The equation utilized in logistical Regression is: p/1-

p = exp (b0 + b1x)

A. DecisionTrees

Decision tree is a tree structure that is within the type of a flow chart. It's used as a technique for classification and prediction with illustration exploitation nodes and internodes. The foundation and internal nodes area unit the take a look at casesThat area unit won't to separate the instances with totally different options. Internal nodes themselves area unit the results of attribute take a look at cases. Leaf nodes denote the category variable. Figure 1. Shows a sample call tree structure.

International Journal of information Mining & information Management method (IJDKP) Vol.5, No.1, Gregorian calendar month 2015 five Figure one. Sample call Tree Structure call tree provides a robust technique for classification and prediction in polygenic disorder designation drawback. Varied callalgorithmsareaunitonthemarkettoclassifythe information, together with ID3, C4.5, C5, J48, CART and CHAID. During this paper, J48 call tree algorithmic rule [10] has been chosen to determine themodel.Everynodeforthechoicetreeisfoundby scheming the very best info gain for all attributes associated if a particular attribute provides an unambiguous end result (explicit classification of sophistication attribute), the branch of this attribute is terminated and target worth is allottedthereto.

IV. FLOWCHART

The following fig. 1 depicts a general methodology or the flow of diabetes prediction system. We propose to predict whether a patient will suffer from diabetes or not in the future. In this process we will first collect a database of diabetes from hospitals or laboratories. Then we will be categorizing the collected data into diabetic patient and non-diabetic patient. We then extract the features from the dataset. From these features we select the attributes which are more responsible for diabetes in a patient. This stage is Feature selection. Later, we pre-process these features, to make it easier to process them and also to increase the accuracy. This dataset is then splatted into training dataset and testing dataset. The train-test split is of the ratio 70:30 which means that 70% of the data willbeourtrainingdataandtherest30%willbeour testing data. With the help of feature selection, we are able to get precise accuracy. This output is then evaluated using performance metrics. Machine learning models are used and the one with thebetter accuracy will be chosen to further train themodel.

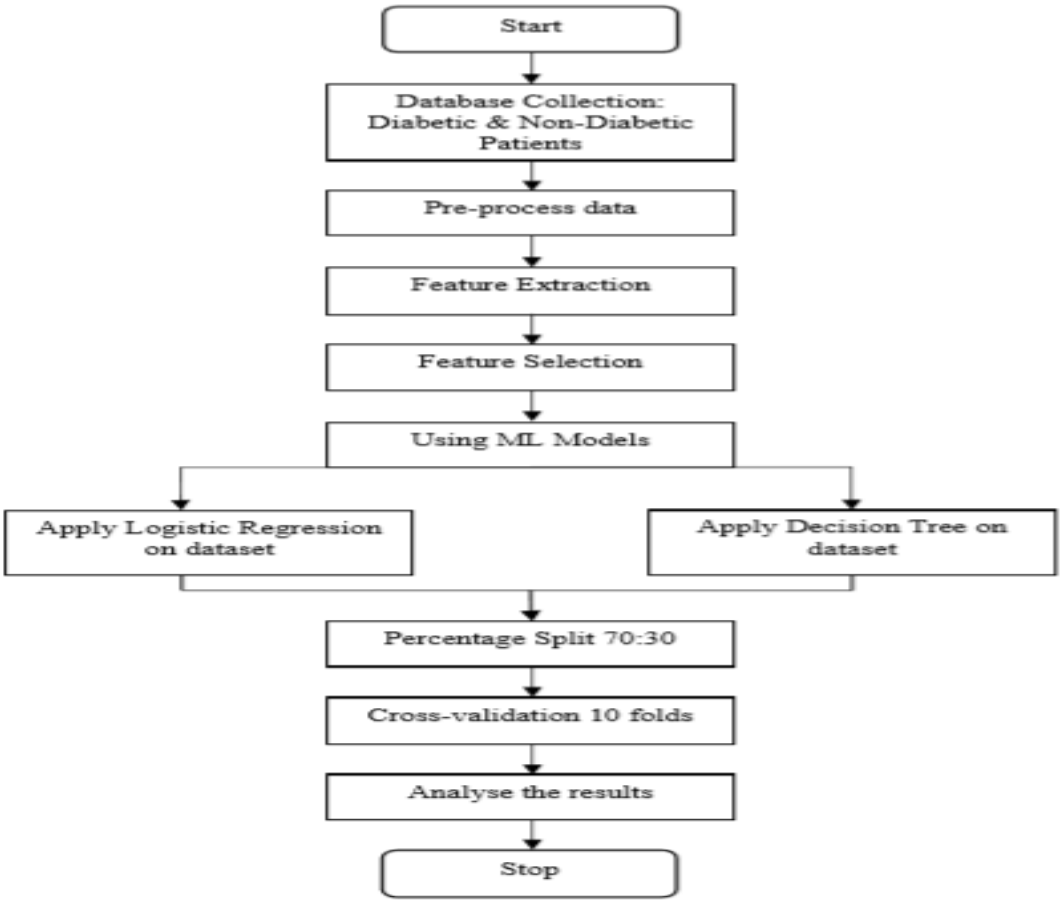


Fig 1. Flowchart of Diabetes Prediction System

IV. DATASET DESCRIPTION AND PRE- PROCESSING

The paper explores the aspect of DecisionTree and Logistic Regression as Data Mining techniques in determining diabetes in women. The classificationtypeofdataminingtechniquehasbeen applied to the Pima Indians Diabetes Dataset of National Institute of Diabetes and Digestive and KidneyDiseases.

TABLE 1

Dataset Description

Dataset	No. of Attributes	No. of Instances
Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney Diseases	8	768

TABLE 2

Attribute Description

Attribute	Re-labeled values
1.Number of times Pregnant	Preg
2.Plasma Glucose concentration	Plas
3.Diastolic blood Pressure (mmHg)	Pres
4.Triceps skin fold thickness(mm)	Skin
5.2-Hour Serum Insulin	Insu
6. Body Mass index (kg/mm^2)	Mass
7.Diabetes Pedigree function	Pedi
8. Age (Years)	Age
9. Class Variable (0 or 1)	Class

4) Age(years)

5) Class Variable (nominal)-determines whether the person is suffering from diabetes or not.

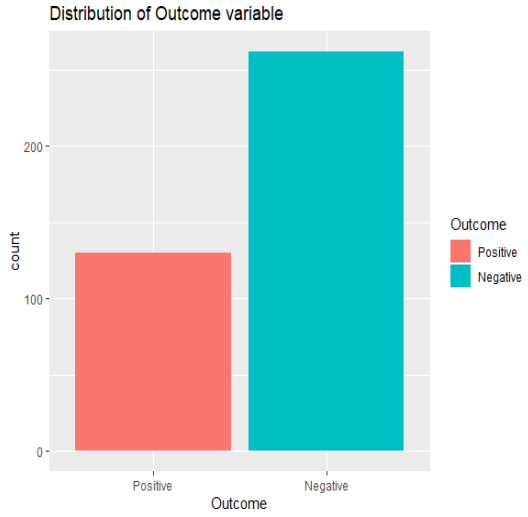


Fig. 2 distribution of the Outcome

VI. IMPLEMENTATION

In this paper two algorithms namely, J48 (decision tree algorithm) and Logistic Regression, have been used to create the model for diagnosis. The dataset was divided into training dataset and testing dataset by the cross-validation techniqueand percentage splittechnique.

10-fold cross validation is used to prepare training and testing dataset. After data pre-processing (CSV format), the Logistic algorithm is employedonthedatasetafterwhichdataaredivided into "tested-positive" or "tested-negative" dependingonthefinalresultofthedecisiontreethat is constructed. The algorithm is asfollows:

A. Algorithm

Input:

- Pima Indians Diabetes Dataset of National Institute of Diabetes and Digestive and Kidney Diseases dataset pre-processed in CSV format.Transformation steps include:
- Replacing missing values,and
- Normalization ofvalues.

The latter makes it easy to use the dataset, as the range of the variables are 0 and 1. The significant attributes obtained after execution are as follows:

1) Plasma glucoseconcentration
2) Body mass index (kg/m2)
3) Diabetes pedigreefunction

Output:

- J48 Decision Tree Predictive Model with leaf node either tested-positive or tested negative and Logistic Regression PredictionResults.

(1) The dataset is pre-processed. The further operations are performed on the datasetare:

(i) Replace Missing Valuesand

(ii) Normalization ofvalues.

(2) Processed dataset is passed through feature selection wherein certain attributes are deleted from thedataset.

(3) The final processed dataset is uploaded in csv file.

(4) The J48 Decision Tree and Logistic Regression algorithm areemployed.

(5) For purposes of the algorithms, Cross-Validation and Percentage Split techniques are applied for modelvalidation.

B. Software ApplicationApproach

With the advent of smartphones, patients are increasingly using mobile technology through automated text messages and various applications, or"apps,"tomonitortheirdiseasestates.In2010,an estimated 10.9 million people > 65 years of age and 215,000people<20yearsofageintheUnitedStates haddiabetes.

In the fast-paced and evolving world of technology, youngerpopulationsareoftenhighlyproficientwith and adaptable to smartphones. As a result, mobile apps for Apple iOS and Android operating systems have overwhelming popularity. More than 500 million Apple iOS and Android devices have been activated since2007.

In fig.3 the system architecture of diabetes prediction system is shown. We have trained diabetes dataset using Machine learning models. This software application is beneficial for both non- diabetic and diabetic patients. While training the model, age is one of the important factors for predicting diabetes. Hence this application can be used by any age group. For the people not suffering from diabetes (based on age, weight and height) can also use this application in order to predict if they will be prone to diabetes in the near future. The people suffering from diabetes can keep track of the factors such as blood pressure, insulin levels, glucose etc. on a daily basis. The application aims to prevent the onset of diabetes by forecasting the disease earlier and the precautions could be taken at an early stage.

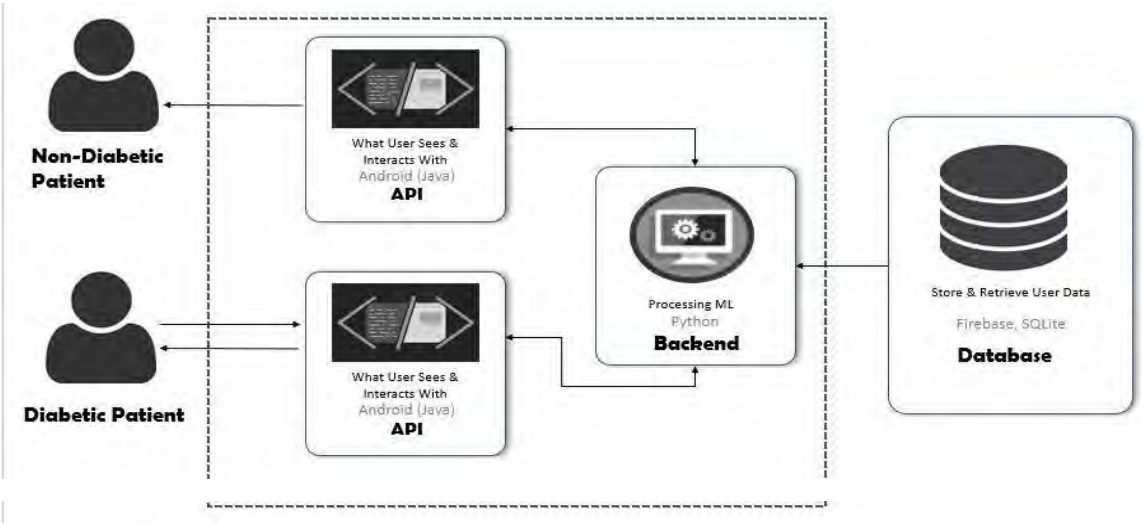


Figure. 3 System Architecture of Diabetes PredictionSystem

VII. RESULT ANDANALYSIS

A. LogisticRegression

1) Based on Percentage Split (70:30)Technique: Since a 70:30 percentage split was applied on the dataset 119 of the instances were used as the test dataset while the rest of were using for training the model. The logistic algorithm gives the following correctness results for the givendataset.

TABLE 3

Confusion Matrix and Statistics

	Prediction	
Actual	0	1
0	71	17
1	8	22

TABLE 4

Performance Measures from Logistic Regression

Accuracy	0.7881
95% CI	(0.7033,0.858)
No Information Rate	0.6695
P-Value[Acc>NIR]	0.003185
Kappa	0.4916
Mcnemar's Test P-Value	0.109599
Sensitivity	0.8987
Specificity	0.5641
Positive Pred Value	0.8068
Negative Pred Value	0.7333
Prevalence	0.6695
Detection Rate	0.6017
Detection Prevalence	0.7458
Balanced Accuracy	0.7314
'Positive' Class	0

4) Specificity (also called the true negative rate) measures the proportion of actual negatives that are correctlyidentified.

5) Mean Absolute Error: average of the absolute error between observed and predictedvalue.

6) Root Mean Squared Error: It measure of the differences between value (Sample and population values)predictedbyamodelor anestimatorandthe values actuallyobserved

7) Relative Absolute Error: It is the ratio of the absolute error of the measurement to the accepted measurement.

2. Terminologies of TestStatistics

1) Kappa Statistic: It is a metric that compares the Observed Accuracy with ExpectedAccuracy.

2) Mcnemar's Test P-Value: McNemar's Test [13] (sometimes also called "within-subjects chi-squared test") is a statistical test for paired nominal data. In context of machine learning models, we can use McNemar's Test to compare the predictiveaccuracy of the twomodels.

3) Sensitivity (also called the true positive rate, the recall, or probability of detection in some fields) measures the proportion of actual positives that are correctlyidentified.

TABLE 5

	Prediction	
Actual	0	1
0	233	5
1	29	74

Confusion Matrix of Logistic Regression

In the table, the values represent following:

a. Numberofcorrectoutcomethattheinstancetested

positive
2. Number of incorrect outcome that the instance testednegative
3. Number of incorrect outcome that the instance testedpositive
4. Number of correct outcome that the instance testednegative

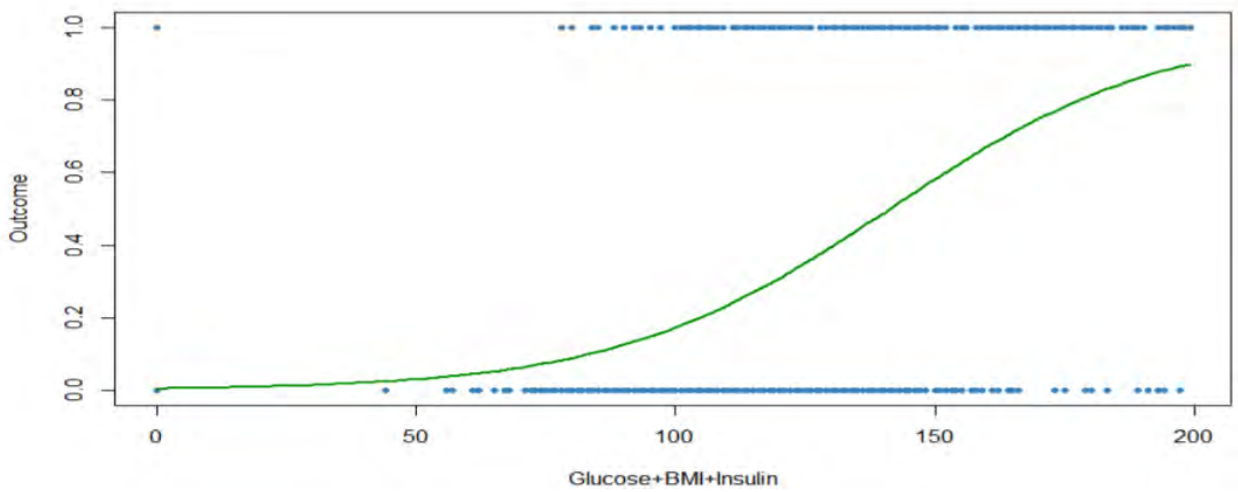


Figure. 4 Logistic Regression Curve

B. DecisionTree

Decision tree J48 implements Quinlan's C4.5 algorithm [5] for generating pruned tree. The tree generated can be used for classification of a patient that has being tested positive or negative for diabetes.Thedataminingtechniqueusesthethought ofinformationgain.Eachattributeofthedataisused to make a decision by splitting the data into smaller orsub-modules.

It examines normalized information gain (difference in entropy) that results from choosingan attribute as a split point. The highest normalized IG is used at the root of the decision tree. This procedure is repeated until the leaf node is created for the tree specifying the class attribute that is chosen. From the results obtained, both themethods have a

comparatively small difference in error rate, though the percentage split of 70:30 for Logistic Regression technique gives the least error rate compared to other two decision tree implementations. The machine learning models are efficient in the diagnosis of diabetes using the percentagesplitof70:30ofthedataset.Adeveloped model for diagnosis of diabetes will require more trainingdataforcreationandtestingwhichwouldbe more beneficial for our softwareapplication.

C. PerformanceAnalysis

Performanceofthemodelisbeingevaluatedon the basis of ROC curve. ROC Curve stands for Receiver Operating Characteristics Curve which is represented below in the fig, 7. The threshold value we took for analyzing the performance is0.5.

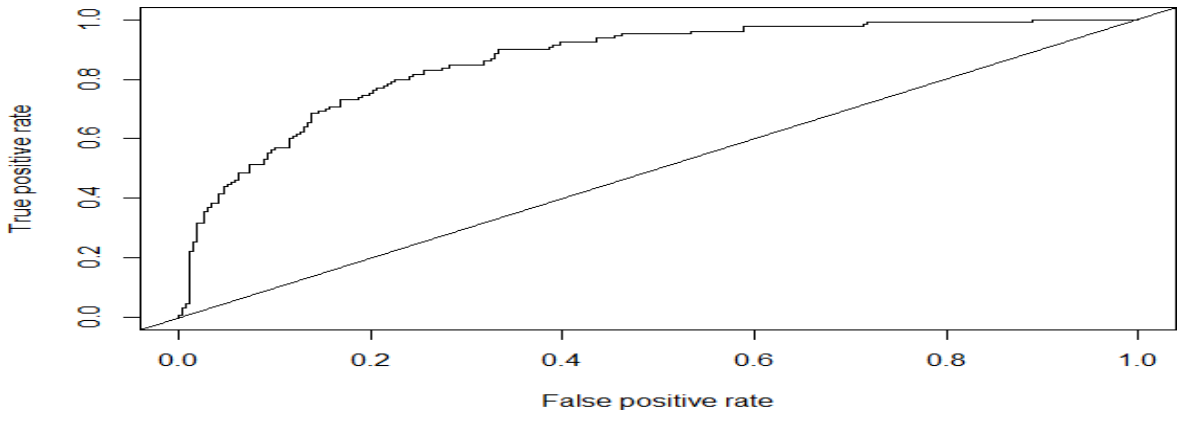


Fig.7 ROC Curve

By using both the supervised learning models, it is known that the Logistic regression model works better on classified or labelled data. The accuracy of the Logistic Regression came out to be 0.7884 whereas that of Decision Tree came out to be 0.6933. In fig. 6 we can easily predict that the Logistic Regression Model gives better performance and less error rate than Decision Tree. This result will be helpful in order to achieve good results for real time datasets.

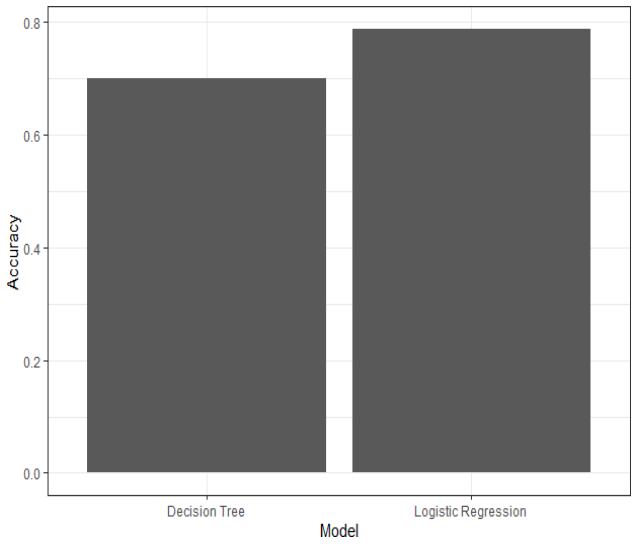


Fig. 6 Comparison of Model Accuracy

VIII. CONCLUSION

The automatic prediction of diabetes is an important real world medical problem. Detection of diabetes in its early stages is the successful key for treatment. This paper shows how the Data Mining techniques are used to model actual diagnosis of diabetes for local and systematic treatment, along with presenting related work in the field. Experimental results show the effectiveness of the projected model. The performance of thetechniques was investigated for the diabetes designation downside. Experimental results demonstrate the acceptableness of the projected model. In future,it's plannedtoassembleadditionaldatafromcompletely different locales everywhere the globe and build a additional precise and general discerning model for diabetes conclusion. Future study can likewise specialize in gathering data from a later period and see new potential prognostic parts to be incorporated. The work may be extended and improved for the automation of diabetesanalysis.

IX. FUTURESCOPE

There is so much to learn in this vast field of datascience.Inourproject,wehaveonlyfocusedon one set of population living in a particular region. Our aim is to expand our work on a city, where we would add more attributes responsible for diabetes. This would be a good estimation for the onset of diabetes. This would alert people and could take precautions beforehand. We aim to prepare a website where people would enter their details and would predict whether they would be prone to such disease called diabetes in the nearfuture.

References

[1] We have referred "www.kaggle.com" for the data set for practicing algorithms. National Diabetes Information Clearinghouse(NDIC),http://diabetes.niddk.nih.gov/dm/pubs/type1and2/#signs

[2] Global Diabetes Community,http://www.diabetes.co.uk/diabetes_care/blood-sugar-level-ranges.htmlJiawei Han and Micheline Kamber, "Data Mining Concepts andTechniques",Morgan Kauffman Publishers, 2001

[3] S. Kumari and A. Singh, "A Data Mining Approach for the Diagnosis of Diabetes Mellitus", Proceedings of Seventh lnternational Conference on Intelligent Systems and Control, 2013, pp. 373-375

[4] Michael Feld, Dr. Michael Kipp, Dr. Alassane Ndiaye and Dr.DominikHeckmann "Weka: Practical machine learning tools and techniques with Java implementations"

[5] McNemar, Quinn, 1947. "Note on the sampling error of the difference between correlated proportions or percentages". Psychometrika. 12 (2):153–15

SENTIMENT ANALYSIS: A DEEP DIVE

Palak Furia , Vaishnavi Gupta , Anand Khandare
Department of Computer Science
Thakur College of Engineering and Technology (TCET)
palakfuria18@gmail.com , guptavaishnavi250@gmail.com , anand.khandare@thakureducation.org

ABSTRACT

Sentiment analysis is the interpretation and classification which categorize the feeling in diverse manner like positive, negative etc. allowing businesses to identify customer sentiment toward products, brands or services in online conversations and feedback. It is one of the most popular research field in natural language processing as it is widely studied in data mining, Web mining, and text mining. In fact, this research has spread beyond computer science to the management sciences and social sciences due to its importance to business world and society as a whole. Consider a sentence, *you're sosmart!* It sounds like a compliment However,consider the same sentence in the following context. *Wow, did you think you're so smart?* Now we are dealing with the same words but different emotions. Here sentiment analysis using NLP becomes complex to achieve accuracy .in this paper we discuss all the different method used for sentiment analysis and its use in real world.

KEYWORDS: -sentiment analysis, machine learning, NLP, social media, lexicon based approach, hybrid approach.

I. INTRODUCTION

Sentiment is an attitude, thought, or judgment which is prompted by feeling. Sentiment analysis [1-8], which is also known as opinion mining, is used to study people's sentiments towards certain entities. Web is a resourceful place in order to find sentiment information. From a user's perspective, people post their own content through various forums such as social media, micro-blogs, or social snetworking sites. From a researcher's perspective, many social media sites release their application programming interfaces (APIs) for data collection which is used by researchers, developers and analyst. For instance, Twitter uses three different versions of APIs available [9], which namely are the REST API, the Search API, and the Streaming API. With the REST API, this helps developers to find statistical data and user data; the Search API helps developers to query specific Twitter content, whereas the Streaming API is used to collect real time twitter content. Moreover, developers mix different APIs to create their own applications. Hence, sentiment analysis have a strong fundament with the support of massive data available online.

However, data online have several flaws that potentially disturb the process of sentiment analysis. The first flaw is that since people are able to freely post content online, the opinion quality cannot be guaranteed [10-12]. The second flaw is that the data online is not always true. The Stanford Sentiment 140 Tweet Corpus [13] is a dataset that has ground truth and is also available publicly. The corpus contains 1.6 million machine-tagged Twitter messages. Each message is tagged based on the emoticons (☺as positive, ☹as negative) discovered inside the message. Data is collected from Amazon [14], between February and April, 2014. The flaws faced have been overcome in the following two ways: First, each product review is inspected before it is be posted a. Second, each review has a rating on it that can be used as the ground truth. The rating is based on a star-scaled system, where 5 stars is the best rating and the worst rating has only 1 star.Social media has become one of the major root of information wrapping a wide scope of topics. It has become of the essential platform for exchanging the information. Social media also refers to the online interactive media resources which are posted on social media, it also advance personal engagement and community curation. A huge amount of multimedia data is generated in all type of social network per minute. With increasing demand of camera –equipped mobile terminals and social network platform (e.g. Twitter, Instagram, Facebook, We chat and Weibo), multimedia content including videos, document, images a main role in fetching information according to people's judgment and sentiments in social networks [35, 50, 94]. Sentiment analysis is no longer just a subject on natural language processing but it also have a connection with pattern recognition, computer vision and other issues in artificial intelligence. A variation of new idea have appeared in the promising area, specifically for visual sentiment analysis.

We will focus on various modes of sentimental analysis. A depth information of its different types of application in real world will be given and how it can help in grasping about various changes taking place in the current trends.

II. LITERATURE SURVEY

Minhoe Hur et al., 2016 [8] suggested a system to predict collection of Box-office depended on Sentiments of movie review. For this, they have used Viewer opinions as input variables. A three machine learning-based algorithms (artificial neural network, regression tree, and support vector regression) were used to get rough relationship between the box-office and its collection predictors.

Aurangzeb Khan, 2011 [19] suggested a rule-based technique. In this technique, SentiWordNet is used to get more specific than a pure lexicon-based technique

for sentiment analysis. The accuracy of the system is 86% at sentence level and 91% at document level.

Mudinas and Zhang, 2012 [22] suggested a hybrid technique which excels better in
performance than the lexicon. It also performs like leaning based technique. Hybrid Techniques is more substantial than lexicon technique and its performance is based on machine learning techniques. The overall accuracy of the system is 82.3%.

Lei Zhang et al., 2010 [18] suggested a ranking and extracting product attribute that was in opinion documents algorithm. Firstly, they had difficulty to determine the reviews given by the users by using a machine. And also to segregate them as positive and negative reviews. In order to overcome this, they used rule mining technique to for extracting product features.

Seven Rill et al., 2014 [13] suggested an application "PoliTwi" which shows an emerging political topics on Twitter by detecting it earlier and it gives the impact on concept-level sentiment analysis. For this, hashtags are used to decide the outcome of election in USA even prior to "Google Trends". Twitter API is used to analyse the collected data by using sentiment analyzing an algorithm

Monu Kumar and Dr. Manju Bala, 2016 [7] suggested that it is not easy to analyse the ample of unstructured data these days which are gathered from different social networking
sites like "Facebook, Wechat, Weibo, Instagram". That's why, they use Hadoop for intelligent analysis and cloud service for storage of big data. Cloud is used for Sentimental Analysis of Twitter.

Martin Wöllmer et al., 2013 [16] suggested a technique to analyze sentiments in Audio – Video context of a YouTube Movie. For this, they use Metacritic database as input to get user reviews. They assess the knowledge-based approach by applying data-based approach in-domain setting as well as in a cross-domain setting.

III. TYPES OF SENTIMENT ANALYSIS

Sentiment analysis presumes different forms, from models that focus on polarity (positive, negative, neutral) to those that detect feelings
and emotions (angry, happy, sad, etc), or even models that identify intentions (e.g. interested v. not interested).

3.1 FINE-GRAINED SENTIMENT ANALYSIS:

In business if polarity precision is important to you then you might consider in expanding your polarity categories to include:

- Very positive Positive
- Neutral Negative
- Very negative

3.2 EMOTION DETECTION:

Emotion detection focus on detecting emotions, like happiness, anger, sadness, and so on. Many emotion detection systems use lexicons (i.e. lists of words and the emotions they convey) or complex machine learning algorithms.

One of the drawbacks of using lexicons is that people express their feeling in different ways. Some words that typically express anger, like bad or kill(e.g. your customer support is killing me) might also express happiness (e.g. you are killing it).

3.3 ASPECT-BASED SENTIMENT ANALYSIS:

While analyzing the sentiment of a text or viewing the product review we always want to know what particular features or aspect do people mention are positive, negative or neutral way. That's where aspect-based sentiment analysis can help, for example in
this text: "The battery life of this smartwatch is
too short", an aspect-based classifier would beable to determine that the sentence expresses a negative opinion about the feature battery life.

3.4 MULTILINGUAL SENTIMENT ANALYSIS

Multilingual sentiment analysis can be difficult. It includes a lots of pre-processing and resources. Some of these resources are available online (e.g. sentiment lexicons), while others need to be created (e.g. translated corpora or noise detection algorithms), but you need to know how to code to use them.

Alternatively, a language in texts can be detected automatically, then train a custom sentiment analysis model to classify texts in the language of your choice.

IV. APPROACHES OF SENTIMENT ANALYSIS

Sentiment analysis uses different Natural Language Processing (NLP) methods and
algorithms, which we'll go over in more detail in this section.

Rule-based systems are used toperform sentiment analysis based on a set of instruction crafted rules.

Automatic systems rely on machinelearning techniques to learn from data.

Hybrid systems combine both rule-based and automatic approaches.

The machine learning-based approach uses a classification technique to categorize the text; it includes two sets of documents: training and a test set. The training set is used for learning to transform characteristics of a document, while the test set is used for inspecting how well the classifier performs.

The **lexicon-based** approach uses sentiment dictionary with opinion words and matches them with the data for determining polarity. There are three different techniques to construct a sentiment lexicon: manual construction, corpus-based methods and dictionary-based methods. The manual construction is the most difficult and time-consuming task. Corpus-based methods can produce opinion words with relatively high accuracy. Finally, in the dictionary-based techniques, the idea is to first collect a small set of opinion words manually with known orientations, and then to grow this set by searching in the WordNet dictionary for their synonyms and antonyms.

Finally, in the **hybrid approach**, the combination of both the machine learning and the lexicon-based approaches has the potential to better the sentiment classification performance.

SENTIMENT CLASSIFICATION APPROACHES		FEATURES/ TECNIQUES	ADVANTAGES AND LIMITATIONS
Machine learning	Bayesian Networks Naive Bayes Classification Maximum Entropy Neural Networks Support Vector Machine	Term presence and frequency Part of speech information Negations Opinion words and phrases	**ADVANTAGES** the ability to adapt and create trained models for specific purposes and contexts **LIMITATIONS** the low applicability to new data because it is necessary the availability of labeled data that could be costly or even prohibitive
Lexicon based	Dictionary based approach Novel Machine Learning Approach Corpus based approach Ensemble Approaches	Manual construction, Corpus-based Dictionary-based	**ADVANTAGES** wider term coverage **LIMITATIONS** finite number of words in the lexicons and the assignation of a fixed sentiment orientation and score to words
Hybrid	Machine learning Lexicon based	Sentiment lexicon constructed using public resources for initial sentiment detection Sentiment words as features in machine learning method	**ADVANTAGES** lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain **LIMITATIONS** noisy reviews

Table 1. comparing different analysis approach

V. WORKING OF SENTIMENT ANALYSIS

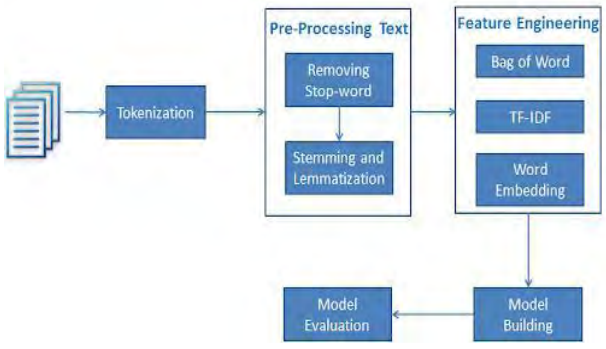


Fig1. Flow chart

The process of sentimental analysis include following steps:
Step1: training and predicting
In the process of training, model (dataset) is made to learn how to associate a particular set of input(text) to the output (tag) correspondingly based on the test sample used for training. Pair of feature vectors and tags are fed into algorithms by feature extractors to generate a model.

In the process of prediction, the feature extractor transforms the unseen text inputs into feature vectors. These feature vectors are fed to the model which generates prediction tags.

Step2: feature extraction
Text classifier in machine learning is of the most important step , transforming the vectorized text helps in finding similar word
representation with similar word meaning.Word embedding also known as word vectors is a new extraction technique, other common technique commonly used is bag-of-wordsakabag-of-ngrams

Step3: algorithm
Classifying the algorithm to be used is generally based on the statistical model like Naïve Bayes,
Logistic Regression, Support Vector Machines, or Neural Networks etc:

- Naïve Bayes: a family of probabilistic algorithms, Bayes's Theorem is used to predict the category of a text.
- Linear Regression: most commonly used algorithm in statistics which user a given set of features (x) to predict the values (y).
- Support Vector Machines: a non-probabilistic model which uses a representation of text sentiments are mapped to distinct regions within that space. New text are assigned similar category they're mapped to
- Deep Learning: a diverse set of algorithms that tries to mimic the human brain, by employing artificial neural networks in processing of data.

VI. ACCURACY AND CHALLENGES OF SENTIMENTAL ANALYSIS

6.1 ACCURACY:

Sentiment analysis is difficult due to things such as subjectivity, tone, lack of context, irony, and sarcasm. Not only the algorithms is a factor that is hard for doing sentimental analysis but also humans.According to Saif et al., human-agreement for Twitter sentiment analysis reaches a 0.655 value of Krippendorff's Alpha, which is a quiet exceptional of agreement but still far from great. According to our experience, custom machine learning models for sentiment analysis can reach 70–80% accuracy level with proper training, and sometimes it can even give higher result depending upon the on the domain and the scope of the problem. If one is interested in sentiment analysis, then they should check out the complete guide which goes through the basics of sentiment analysis, how does it work, it's challenges and applications, and share some useful resources.

6.2CHALLENEGES:

6.2.1SUBJECTIVITY AND TONE:

The detection of subjective and objective texts is just as essential as analysing their tone. However, so called *objective* texts does not contain explicit sentiments.

6.2.2 IRONY AND SARCASM:

People express their negative sentiments using positive words, which can be difficult for machines to detect the context without having a thorough understanding of the situation in which a feeling was expressed. This is what is called as irony and sarcasm.
For example, look at some possible answers to the question, *Did you enjoy your shopping experience with us?*
Yeah, sure. So smooth!
Not one, but many!

6.2.3 COMPARISONS:

Comparisons in sentiment analysis is another challenge which is worth tackling. Look at the texts below:

This product is second to none.
This is better than older tools.
This is much better than nothing.

6.2.4 EMOJIS:

According to Guibon et al. there are two types of emojis. *Western emojis* (e.g. :D) are encoded in only one or two characters, whereas *Easternemojis* (e.g. ¯_(ツ)_/¯) are a longercombination of characters of a vertical nature. Emojis play an important role in the sentiment of texts, particularly in tweets.

6.2.5 DEFINING NEUTRAL:

To define neutral is one more challenge to tackle in order to get accurate sentiment analysis. As in all problems, defining your categories -and, in this case, the *neutral* tag- is one of the almost important parts of the problem. The *neutral*, *positive*, or *negative* does matter when you trainsentiment analysis models. A good problem definition is a must as tagging data requires the tagging criteria.

VII. TOOLS USED FOR SENTIMENTAL ANALYSIS:

Sentimental analysis is a deep topic and not all beginners are comfortable using languages like R, python etc, there are number of blogs and forums available with tutorials of how to analyse but still it is easier to directly use application.

TOOLS FOR SENTIMENT ANALYSIS	TECNIQUES USED BY TOOLS
EMOTICONS	Emoticons contained in the text
LIWC	Dictionary and sentiment classified categories
SentiStrengh	LIEC dictionary with new features to strength and weak sentiments
Senti WordNet	Lexical dictionary and scores obtained by semi-machine learning approaches
SenticNet	Natural language processing approach for inferring the polarity at semantic level
Happiness Index	Affective Norms for English Words (ANEW) and scores
	for evaluating happiness in the text
AFINN	Affective Norms for English Words (ANEW) but more focused on the language used in microblogging platforms.
PANAS-t	Eleven-sentiment psychometric scale
Sentiment140	API that allows classifying tweets to polarity classes positive, negative and neutral.
NRC	Large set of human-provided words with their emotional tags.
EWGA	Entropy-weighted genetic algorithm
FRN	Feature relation network considering syntactic n-gram relations

Table 2. tools

VIII. APPLICATIONS

8.1 SOCIAL MEDIA MONITORING:

- It can be used to Analyse tweets and/or facebook posts to detect audience sentiment over a particular topic over a period of time.
- Control social media mentions of your brand
- Automatic route finding in social media to mention same team members
- Get deep insight of current social media trends.

8.2 BRAND MONITORING:

Not only brands have large amount of information available on social media, but all across internet. Instead of focusing specifically only on social media, they also use information from news, blogs, forums etc to analyse the brand status in the society and implement certain changes accordingly after studying the feedbacks.

8.3 MARKET RESEARCH:

sentiment analysis promotes all kinds of market research and competitive analysis. It helps in exploring and anticipating future trends and and also promotes competitive edge.

SENTIMENT ANALYSIS APPLICATIONS
BUSINESS
Consumers voice
Brand reputation
Online advertising: Blogger Centric Contextual Advertising Dissatisfaction oriented online advertising
On-line commerce
POLITIC
Voting advise applications
Clarification of politicians' positions
PUBLIC ACTIONS
Real-world events monitoring
Legal matters "blawgs"
Policy or government-regulation proposals
Intelligent transportation systems

Table 3, applications of sentimental analysis

IX. CONCLUSION

Sentiment analysis is a concept that can be applied to number of aspects from political analytics, product analytic, brand monitoring, market research etc. by incorporating this in the
current system we are able to increase the work with useful ends and more accuracy. It has moved beyond just being a topic of interest, it enables us to glean a new insight perspective, so that we can do work better and more productively.

In short, ample of efforts is been devoted to sentiment extraction of social media text. Although considerable amount of progress is been made in visual sentiment analysis but multimodal sentiment is still in its inception. Audio-only sentiment analysis is not usual because audio data is not self-dependent in social

networks, but they are incorporated with the videos. Therefore, the combination of audio-visual information is made for effective use for multimodal sentiment.

References

[1] Kim S-M, Hovy E (2004) Determining the sentiment of opinions In: Proceedings of the
20th international conference on Computational Linguistics, page 1367.. Association for Computational Linguistics, Stroudsburg, PA, USA.Return to ref 1 in article

[2] Liu B (2010) Sentiment analysis and subjectivity In: Handbook of Natural Language Processing, Second Edition.. Taylor and Francis Group, Boca.

[3] Liu B, Hu M, Cheng J (2005) Opinion observer: Analyzing and comparing opinions on the web In: Proceedings of the 14th International Conference on World Wide Web, WWW '05, 342–351.. ACM, New York, NY, USA.

[4]Pak A, Paroubek P (2010) Twitter as acorpus for sentiment analysis and opinion mining In: Proceedings of the Seventh conference on International Language Resources and Evaluation.. European Languages Resources Association, Valletta, Malta.

[5] Pang B, Lee L (2004) A sentimental
education: Sentiment analysis using subjectivity summarization based on minimum cuts In: Proceedings of the 42Nd Annual Meeting on Association for
ComputationalLinguistics,ACL'04..
Association for Computational Linguistics, Stroudsburg, PA, USA.

Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr2(1-2): 1–135.

Turney PD (2002) Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, 417–424.. Association for Computational Linguistics, Stroudsburg, PA, USA.
Whitelaw C, Garg N, Argamon S (2005) Using appraisal groups for sentiment analysis In: Proceedings of the 14th ACM International Conference on Information and Knowledge
Management, CIKM '05, 625–631.. ACM, New York, NY, USA.Return to ref 8 in article

[9] Twitter (2014) Twitter apis. https://dev.twitter.com/start.
Liu B (2014) The science of detecting fake reviews. http://content26.com/blog/bing-liu-the-science-of-detecting-fake-reviews/.

Jindal N, Liu B (2008) Opinion spam and analysis In: Proceedings of the 2008 International Conference on, Web Search and

Data Mining, WSDM '08, 219–230.. ACM, New York, NY, USA.

Mukherjee A, Liu B, Glance N (2012) Spotting fake reviewer groups in consumer reviews In: Proceedings of the 21st, International Conference on World Wide Web, WWW '12, 191–200.. ACM, New York, NY, USA.

[13] Stanford (2014) Sentiment

http://www.sentiment140.com/.Returnto ref 13 in article

[14] www.amazon.com.

Movie Recommendation System Using R

Kartik Rawool, Anurag Tiwari & Suraj Jha

Thakur College of Engineering and Technology Mumbai, India

Abstract— **Recommendation systems have become ubiquitous today. With the advancements undertaken in the movie's domain in the past few decades, it becomes an arduous task to choose movies based on personalized preferences. Therefore, the need for a reliable system that recommends moviesto users becomes inevitable. Such a system takes into consideration a user's past viewing habits, tastes, personal fondness over a genre, actor, director, and latest trends in order to recommend movies to the person. In this paper, we highlight a user based collaborative filtering system that takes an input of three movies from a person based upon their preferences and recommends a list of ten movies which resemble the user's original choices. The filtering and recommendation is generated based upon ratings of several different users on many different movies. The system significantly enhances the user interface and gives a friendlier interface to interact with the application. It also plots and visualizes the data in order to build a more comprehensive understanding of the vast inclusive set of data working behind the application. In an attempt to connect and find similarities between different users, we explored the different similarity techniques which included Pearson, Jacquard and Cosine. Our application recommends ten movies to users based upon historical ratings and similarity between other users. In the final outcome, an interface in Shiny package in R studio was implemented through which the plots and movies output wasproduced.**

Keywords— Movie Recommender, User based Collaborative filtering, Recommendation model

I. INTRODUCTION

In today's rapid world, information overload has become a reality. Filtering out relevant content has become a hot research topic which has gained a lot of attention. [1] Movies are a pivotal source of entertainment. In presence of such diverse and varied content streaming on the internet on a seamless basis today, it becomes difficult to filter out content in a personalized format. On the other end, businesses and industries producing this content face problematic issues in reaching to the right kind of audience, in the absence of a marketing strategy. In order to ease out the selection of movies from a set of perplexing choices, the need and demand of a reliable recommendation engine becomes inevitable. It not only simplifies the selection mechanism for the users but also helps the businesses, industries as well as the creators to reach the right kind of audience. Providing related content which resembles a user's choices out of relevant and irrelevant collection of movies pertaining to his taste is the precise task of a recommendationengine.

Recommendation engines have become increasingly popular as they have been implemented at many different online platforms. The content of such platforms varies from movies, music, books and videos, to friends and stories on social media platforms, to products on e- commerce websites, to people on professional and dating websites, to search results returned on Google. [2

These systems are used to collect information from the user and to give him an insight into the related content according to his specific preferences. A wide variety of approaches have been used to provide recommendations. These incorporate collaborative filtering, content-based filtering and a hybrid approach. Different Algorithms and approaches are present to provide recommendations that use rating or content information; however collaborative filtering and content-based method suffer from the same limitations. Several researchers have tried to overcome these limitations by combining both collaborative filtering and content-based method and producing a hybrid approach that combines ratings as well as content information. Recommendation system will always remain an active search area for researchers.[3]

The table gives us conclusive data about how useful the recommendation engines are for the businesses as well. [4]

Netflix	2/3rd of the movies watched are recommended
Google News	recommendations generate 38% more click-troughs
Amazon	35% sales from recommendations
Choicestream	28% of the people would buy more music if they found what they liked

Table1. Companies benefit through recommendation system

II. RELATEDWORK

Inthissystem,wehadimplementedamovierecommender model in accordance with which we had referred several websites and implemented the concept in a very systematic and simplified manner. We had collected the required dataset from a website, titled group lens [5]. In order to process that data and apply various models for visualizing the data we underwent a pre-processing techniquewheretherelevantdatacontentwasfilteredout. Inabidtokeepitsimpleweused100,000ratingsand8570 movies with 706 users[6].

In this website we discovered that in a content-based model, we needed more specific and precise information about our user and therefore in a content-based model it was necessary to have a large dataset which in turn would make the system a bit more complex and inefficient. Hence, we make use of the collaborative learning model for recommending movies to a user since it is based on the similarity of the users who have the same preferences as that of another user [6].In this website it used genres for recommending movies. In order to do so, it first converted the genre into a binary matrix in a 0 and 1 format which was arising some number of limitations. So, we made the required changes and modified the model by using only the ratings parameter instead of genre which made the process more simplified. For creating the rating matrix we are using recommender lab function which is available in reshape2 package in R. We then convert itin the wide format rather than long format by cast () function which is in reshape2 package in R. It is also efficient to work with graphics packages lattice and ggplot2, as it makes it easy for creating plots to render comparisons of various data[7].

Now for analyzing the large amount of data it becomes tedious for computing with a greater number of variables, so we used Principal Component Analysis (PCA) for dimensionality reduction. It is used for converting the features i.e. the dimensions. This means restricting 3D to 2D or else 2D to 3D. For example, we are having 2 variables a and b, then we split into one more variable which implies that now, we are having 3 variables a, b, and c. Also, we add one more dimension to it. The function rind () is used to combine vectors/matrix since we are using it in rating mat and user select [8].

Now we implemented the technique of visualization of the dataset and by using knn algorithm to the rating matrix, we found out the further steps in our model to be continued [9].

III. PROPOSEDMETHODOLOGY

Architecture

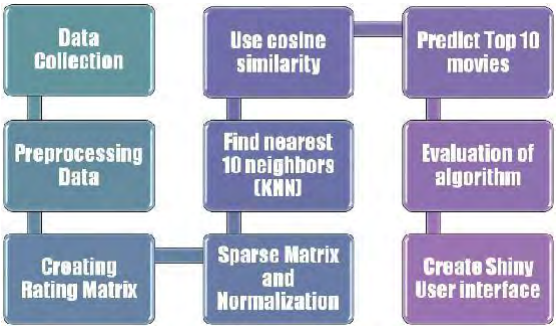


Fig 1: System Architecture

From the datasets, the ratings dataset was converted from the given data frame into a matrix data structure. This was done in order to apply our model later which required that the datasetshouldbepresentinmatrixformat.ThisMatrixisthen converted into a sparse matrix which nullified the ratings whose data was absent andmissing. We have a total of 100K ratings which include about 600+ users. We know that all the users have different opinions and perspectives. Also, each person rates his or her movies in a different way i.e. some users are stringent and critical while rating, whereas some are quite lenient. So, to bring everyone on the same ground we normalize the matrix. This weights the ratings about zero, which helps to compare different users and also gets the proper ratings of movies. We can see the normalized ratings being plotted in the following histogram. From the histogram we can infer that the greatest number of ratings areaverage.

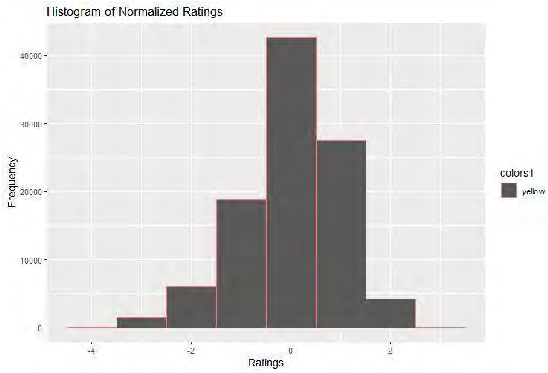


Figure 2: Histogram of Normalized Ratings

We further visualize our dataset using the
Package available in R studio and the plotted results are in the form of histograms and graphs which allow us to compare and give a deeper insight into the dataset which we are dealing with.

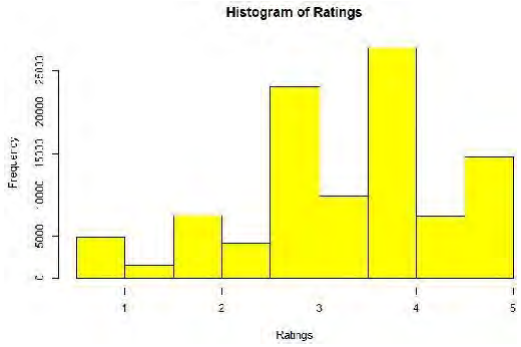


Figure 3: Histogram of Ratings

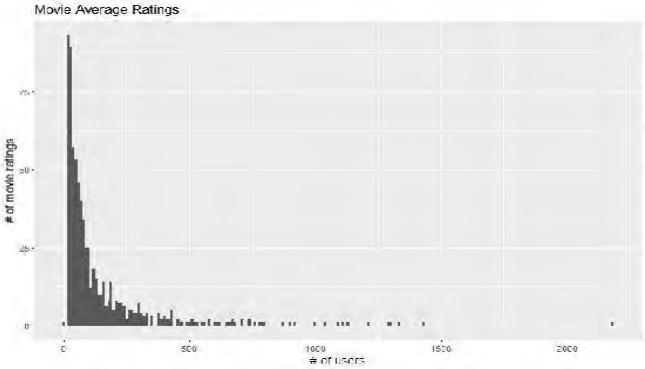


Figure 4: Average rating of movies

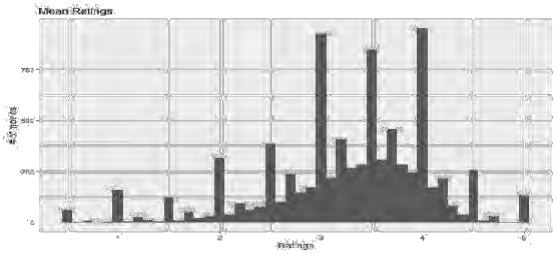
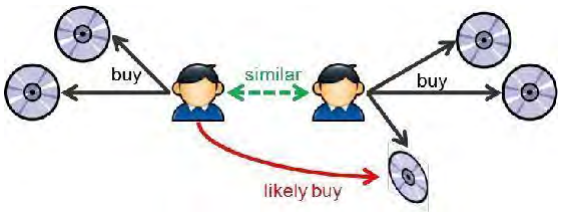


Figure 5: Mean distribution of ratings Once our ratings dataset is complete and ready for processing, we move on to the recommendation model. In this model we are going implement User Based Collaborative Filtering. For this filtering we need tofind users from the dataset who are similar to the person who is using our model. To find the users who have the same preferences, we use KNN algorithm. Through this algorithm we get n nearest neighbors. To get the most similar users we use and implement the similarity techniques. There are 3 types of similarity techniques available: 1. Jacquard 2. Cosine 3. Centered cosine or Pearson similarity. For our model we went ahead with cosine similarity. Higher the cosine similarity better is theselection.

Figure 6: User Based Collaborative Filtering System [10]

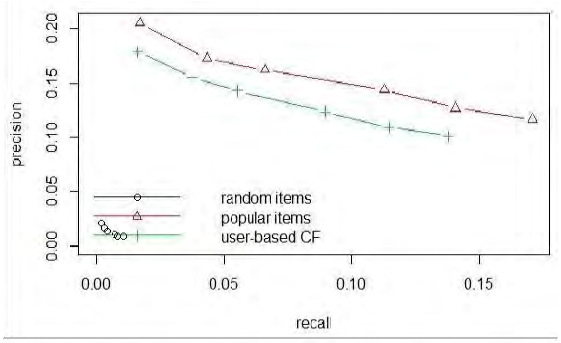


WewillbeusingRprogrammingtoimplementthemodel. So, we used the recommender lab package to use the UBCF model. With this model 10 movies are recommended to the user upon selection of three movies bytheuser.Therearevariouswaystorecommendamovie which include item based, collaborative, randomly or on popular basis but, when we performed the evaluation on the various models, and visualized it on ROC and precision-recall curve, we concluded after observing the curves that the UBCF and Popular are the best based on Area Under Curve. However, we are aiming to recommend on the basis of user's preferences. Hence, we narrowed down on the UBCFmodel.

$$\text{TPR (sensitivity)} = \frac{TP}{TP+FN}$$

$$\text{FPR (1-specificity)} = \frac{FP}{TN+FP}$$

Figure 7: ROC CURVE (RECEIVER OPERATING CHARACTERISTIC)



$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Figure 8: PRECISION RECALL CURVE

The following figure highlights and encompasses the algorithm that was undertaken in order to perform the User Based Collaborative Filtering on our model wherein each stage points to a pertinent step which was followed in the course of the entire implementation.

Figure 9: Working of the User Based Collaborative Filtering

After building the model we applied and implementedthe model in the Shiny web app available in the R Studio. Through this app we accept 3 movie preferences from the user and then display the 10 recommendations using the above model. Also, we display the average rating of the 3 movies out of 5. We also used the shiny app to display all the visualization results. Through this Shiny app wecan

Deploy the app on the web and then the app can be accessed by any user on the internet. This is all possible due to the Shinyapp.io website.

Figure 10: Shiny User Interface

IV. IMPLEMENTATION

Software

The following code is the complete implementation which includes the server and user interface code in R studio.

```
#server.R
library(shiny)
library(ggplot2)
movies <- read.csv("E:/ANURAG/DATA SCIENCE/Final PROJECT/movies.csv", header = TRUE, stringsAsFactors=FALSE)
movies <- movies[with(movies, order(title)), ]
ratings <- read.csv("E:/ANURAG/DATA SCIENCE/Final PROJECT/ratings100K.csv", header = TRUE)
colors = c("red", "yellow", "green", "violet", "orange","blue", "pink", "cyan")
colors1 = c("yellow")
colors2 = c("red")

shinyServer(function(input, output) {

 # Text for the 3 boxes showing average scores

  formulaText1 <- reactive({
    paste(input$select)
  })
  formulaText2 <- reactive({
    paste(input$select2)
  })
  formulaText3 <- reactive({
    paste(input$select3)
  })

  output$movie1 <- renderText({
    formulaText1()
  })
  output$movie2 <- renderText({
    formulaText2()
  })
  output$movie3 <- renderText({
    formulaText3()
  })# Table containing recommendations
  output$table <- renderTable({

    # Filter for based on genre of selected movies to enhance recommendations
    cat1 <- subset(movies, title==input$select)

    cat2 <- subset(movies, title==input$select2)
    cat3 <- subset(movies, title==input$select3)

    # If genre contains 'Sci-Fi' then return sci-fi movies
    # If genre contains 'Children' then return children movies
    if (grepl("Sci-Fi", cat1$genres) | grepl("Sci-Fi", cat2$genres) | grepl("Sci-Fi", cat3$genres)) {
      movies2 <- (movies[grepl("Sci-Fi", movies$genres) , ])
    } else if (grepl("Children", cat1$genres) | grepl("Children", cat2$genres) | grepl("Children", cat3$genres)) {
      movies2 <- movies[grepl("Children", movies$genres), ]
    } else {
      movies2 <- movies[grepl(cat1$genre1, movies$genres)
                        | grepl(cat2$genre1, movies$genres)
                        | grepl(cat3$genre1, movies$genres), ]
    }

    movie_recommendation <- function(input,input2,input3){
      row_num <- which(movies2[,3] == input)

      row_num2 <- which(movies2[,3] == input2
      row_num3 <- which(movies2[,3] == input3
      userSelect <- matrix(NA,length(unique(ratings$movieId)))
      userSelect[row_num] <- 5 #hard code first selection to rating 5
      userSelect[row_num2] <- 4 #hard code second selection to rating 4
      userSelect[row_num3] <- 4 #hard code third selection to rating 4
      userSelect <- t(userSelect)

      ratingmat <- dcast(ratings, userId~movieId, value.var = "rating", na.rm=FALSE)
      ratingmat <- ratingmat[,-1]
      colnames(userSelect) <- colnames(ratingmat)
      ratingmat2 <- rbind(userSelect,ratingmat)

      ratingmat2 <- as.matrix(ratingmat2)

      #Convert rating matrix into a sparse matrix
      ratingmat2 <- as(ratingmat2, "realRatingMatrix")

      #Data Visualization
      output$plot1<-renderPlot(qplot(getRatings(ratingmat2),binwi
```

```
dth=1,col=colors2,main="Histogram of Ratings",xlab="Ratings",ylab = "Frequency"))
        output$plot2<-renderPlot(image(ratingmat2,main="Raw Ratings"))
        rating_norm <-normalize(ratingmat2)
        output$plot3<-renderPlot(image(rating_norm,main="Normalized Ratings"))
        output$plot4<-renderPlot(hist(getRatings(ratingmat2),binwidth=1,col= "yellow",breaks=15,main="Histogram of Ratings",xlab="Ratings",ylab = "Frequency"))
        output$plot5<-renderPlot(qplot(getRatings(rating_norm),binwidth=1,col=colors1, main="Histogram of Normalized Ratings",xlab="Ratings",ylab = "Frequency"))
        output$plot6<-renderPlot(qplot(rowCounts(ratingmat2),binwidth=10,main="Movie Average Ratings",xlab="# of users",ylab = "# of movie ratings"))
        output$plot7<-renderPlot(qplot(colMeans(ratingmat2),binwidth=.1,main="Mean Ratings",xlab="Ratings",ylab = "# of movies"))


        #Create Recommender Model
        recommender_model <- Recommender(ratingmat2, method = "UBCF",param=list(method="Cosine",nn=30))
        recom <- predict(recommender_model, ratingmat2[1], n=30)
        recom_list <- as(recom, "list")
        recom_result <- data.frame(matrix(NA,30))
        recom_result[1:30,1] <- movies2[as.integer(recom_list[[1]][1:30]),3]
        recom_result <- data.frame(na.omit(recom_result[order(order(recom_result))],]))
        recom_result <- data.frame(recom_result[1:10,])
        colnames(recom_result) <- "User-Based Collaborative Filtering Recommended Titles"
        return(recom_result)
      }

      movie_recommendation(input$select, input$select2, input$select3)

    })


    movie.ratings <- merge(ratings, movies)
    output$tableRatings1 <- renderValueBox({
      movie.avg1 <- summarise(subset(movie.ratings, title==input$select),
                              Average_Rating1 = mean(rating, na.rm = TRUE))
      valueBox(
        value = format(movie.avg1, digits = 3),
        subtitle = input$select,
        icon = if (movie.avg1 >= 3) icon("thumbs-up") else icon("thumbs-down"),
        color = if (movie.avg1 >= 3) "aqua" else "red"
      )

    })


    movie.ratings <- merge(ratings, movies)
    output$tableRatings2 <- renderValueBox({
      movie.avg2 <- summarise(subset(movie.ratings, title==input$select2),
                              Average_Rating = mean(rating, na.rm = TRUE))
      valueBox(
        value = format(movie.avg2, digits = 3),
        subtitle = input$select2,
        icon = if (movie.avg2 >= 3) icon("thumbs-up") else icon("thumbs-down"),
        color = if (movie.avg2 >= 3) "aqua" else "red"
      )
    })

    movie.ratings <- merge(ratings, movies)
    output$tableRatings3 <- renderValueBox({
      movie.avg3 <- summarise(subset(movie.ratings, title==input$select3),
                Average_Rating = mean(rating, na.rm = TRUE))
      valueBox(
        value = format(movie.avg3, digits = 3),
        subtitle = input$select3,
        icon = if (movie.avg3 >= 3) icon("thumbs-up") else icon("thumbs-down"),
        color = if (movie.avg3 >= 3) "aqua" else "red"
      )
    })


    # Generate a table summarizing each players stats
    output$myTable <- renderDataTable({
      movies[c("title", "genres")]
    })

}
)
## ui.R
library(shiny)
library(shinydashboard)
library(proxy)
library(recommenderlab)
library(reshape2)
library(plyr)
library(dplyr)
library(DT)
library(RCurl)

movies <- read.csv("E:/ANURAG/DATASCIENCE/Final PROJECT/movies.csv", header = TRUE, stringsAsFactors=FALSE)
movies <- movies[with(movies, order(title)), ]

```

```
ratings <- read.csv("E:/ANURAG/DATA SCIENCE/Final PROJECT/ratings100K.csv", heade r = TRUE)


shinyUI(dashboardPage(skin="purple",
                      dashboardHeader(title = "Movie Recommender System using R"),
                      dashboardSidebar(
                        sidebarMenu(
                          menuItem("Movies", tabName = "movies", icon = icon("star-o")),
                          menuItem("Plot", tabName = "plot", icon = icon("far fa-chart-bar")),
                          menuItem(
                            list(

                              selectInput("select", label = h5("Select 3 Movies That You Like"),
                                          choices = as.character(movies$title[1:length(unique(movies$movieId))]),
                                          selectize = FALSE,
                                          selected = "Shawshank Redemption, The (1994)"),
                              selectInput("select2", label = NA,
                                          choices = as.character(movies$title[1:length(unique(movies$movieId))]),
                                          selectize = FALSE,
                                          selected = "Forrest Gump (1994)"),
                              selectInput("select3", label = NA,
                                          choices = as.character(movies$title[1:length(unique(movies$movieId))]),
                                          selectize = FALSE,
                                          selected = "Silence of the Lambs, The (1991)"),
                              submitButton("Submit")
                            )
                          )
                          )
                      ),

                      dashboardBody(
                        tags$head(
                          tags$style(type="text/css", "select { max-width: 360px; }"),
                          tags$style(type="text/css", ".span4 { max-width: 360px; }"),
                          tags$style(type="text/css", ".well { max-width: 360px; }")
                        ),

                        tabItems(
                          tabItem(tabName = "movies",
                                  fluidRow(
                                    box(
                                      width = 6, status = "info", solidHead = TRUE,
                                      title = "Other Movies You Might Like",
                                      tableOutput("table")),
                                    valueBoxOutput("tableRatings1"),
                                    valueBoxOutput("tableRatings2"),
                                    valueBoxOutput("tableRatings3"),
                                    HTML('<br/>')

                                  )
                          ),
                          tabItem(tabName = "plot",
                                  fluidRow(
                                    box(title = "Raw Ratings",background = "black", width=6, collapsible = TRUE,
                                        plotOutput("plot2")
                                    ),
                                    box(title = "Normalized Ratings",background = "black", width=6, collapsible = TRUE,
                                        plotOutput("plot3")
                                    ),
                                    box(title = "Histogram of Ratings",background = "black", width=6, collapsible = TRUE,
                                        plotOutput("plot1")

                                    ),
                                    box(title = "Histogram of Ratings",background = "black", width=6, collapsible = TRUE,
                                        plotOutput("plot4")
                                    ),
                                    box(title = "Histogram of Normalized Ratings",background = "black", width=6, collapsible = TRUE,
                                        plotOutput("plot5")
                                    ),
                                    box(title = "Movie Average Ratings",background = "black", width=6, collapsible = TRUE,
                                        plotOutput("plot6")
                                    ),
                                    box(title = "Mean Ratings",background = "black", width=6, collapsible = TRUE,
                                        plotOutput("plot7")
                                    )
                                  ))
                        )
                      )
                      )
)

library(shiny)

ui <- fluidPage(


server <- function(input, output, session) {

}
)
```

V. RESULTS AND DISCUSSIONS

In this recommender model we found that for given input of three movies by user, a list of 10 similar movies which suit the tastes of the user are displayed. The main aim of the movie recommender system was to recommend movies to users according to different users rating on a particular movie. We had a set of users with their previous rating on a set of movies, and we predicted the rating that the user can give to a movie they have not previously rated.

Users of movie rentals prefer a movie recommendation that will match their 'taste' in movies and thus wish to invest their time into movies which are preferred by them. The movie rental businesses are eager to use analytics to solve this problem and gain revenues. Using our Movie Recommender System, a movie rental website can identify which movie the user will like the most and recommend similar titles which will increase their business and build customer loyalty. Collaborative Filtering approach was used in our project whereby the similarity of the ratings of the new user with that of the ratings of the other users were identified and recommendations were generated. With User Based Collaborative Filtering (UBCF) approach, the Cosine similarity of new user with other users was calculated and movies that users in the clusters liked were identified. The recommendation matched the similarity of ratings given as input.

VI. CONCLUSION AND FUTURE SCOPE

Through this project we were able to recommend movies to a user by receiving the user's preferences. Here we used based collaborative filtering for recommendation. We analysed different filtering models such as Item based Collaborative filtering, Popular and Random models. By plotting Receiver Operating Characteristic and Precision Recall Curve, we observed the AUC (Area under Curve) to conclude that UBCF is the best model to be used. For implementation of this model in R programming we took help of the Recommender Lab present. This package includes various functions to use the model we desired. For providing our users with proper experience we implemented a web app using Shiny in R Studio.

References

[1] Peng, Xiao, Shao Liangshan, and Li Xiuranm, "Improved Collaborative Filtering Algorithm in the Research and Application of Personalized Movie Recommendations", 2013 Fourth International Conference on Intelligent Systems Design and Engineering Applications, 2013.

[2] Movie Recommender System Project Report, Prateek Sappadla, Yash Sadhwani, Pranit Arora

[3] A Review Paper on Machine Learning Based Recommendation System, Bhumika Bhatt, Prof. Premal J Patel, Prof. Hetal Gaudani, 2014 IJEDR | Volume 2, Issue 4 | ISSN: 2321-9939

[4] A Movie Recommender System: MOVREC, Manoj Kumar, D.K. Yadav, Ankur Singh, Vijay Kr. Gupta, International Journal of Computer Applications (0975 – 8887) Volume 124 – No.3, August 2015

[5] Grouplens website, Datasets on Movies, https://grouplens.org/datasets/movielens/

[6] Building a Movie recommendation engine with R, WordPress, https://muffynomster.wordpress.com/2015/06/07/buil ding-a-movie-recommendation-engine-with-r/

6/7/2015

[7] Principal Component Analysis, Towards Data Science, https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c

[8] Rbind in R example vector data frame fill columns, Statistical Programminghttps://statistical-programming.com/rbind-in-r-example-vector-data-frame-fill-columns/

[9] K nearest Neighbour Classifier Implementation R scratch, data aspirant, http://dataaspirant.com/2017/01/02/k-nearest-neighbor-classifier-implementation-r-scratch/ 1/2/2017

[10] Introduction to Recommender system using collaborative filtering, Towards Data Science, https://towardsdatascience.com/intro-to-recommender-system-collaborative-filtering-64a238194a26

House Price Prediction Using R

Aayushi Sarda, Aunshuman Saha,Sarandeep Singh
Department of Computer Engineering
Thakur college of Engineering and Technology

Abstract— **This paper explores the question of how house are affected by housing characteristics (both internaland external). House prices increase each year, so there is a need for a system to predict pricing in the future. House price prediction favors the researcher to determine the sale price of a house and can help the customer to orchestrate the right time to purchase a house. The price of a home is usually based on three factors: location, product and timing. This study aims to predict house prices based on numerous factors. In normal case scenario, housing price index denotes the summarized price changes of residential estate houses. For a single familypriceprediction,itrequiresmoreaccuratemethod based on location, house kind, size, build year, local amenity, and other factors which could affect the house demand and supply. The housing market is one of the most competitive in terms of pricing and it tends to vary significantly based on varied factors, so it is one of the main areas to implement the process of machine learning tooptimize and predict the prices with the utmost accuracy. Thereforeinthisproject,weanalyzethevariousessential features while predicting housing prices with considerableaccuracy.**

Keywords—Machine Learning, Linear Regression, Decision tree, Multivariate regression, Data cleaning, R Shiny

I. INTRODUCTION

Predicting prices has been a contest for many researchers. Numerous engineers have tried to use various machine learning paradigms to form a model that can accurately calculate prices with only a minimum margin for error.

House Price Prediction is one such problem. These models take into account obvious parameters such as area of the house in squarefeet, no. of bedrooms, garage area, etc. Each of the features in our study is given a certain weight and that determines how important that feature is towards our desired prediction. This is called feature engineering. We have hundreds of different features to choose from, but one of the drawbacks of having a large number of parameters included is the heavy computations that result in making the regression model, and computing the prediction. We haveused the Iowa Housing Dataset for our project. Here, we have defined linear model data using decision tree model and multivariate model, using several features as its input, hence calculated the root mean squared error (RMS value) for the model. This project thus directs us to the application of machine learning algorithms in addition to other techniques to optimize the anticipated result.

II. METHODOLOGY

A. DATACOLLECTION:

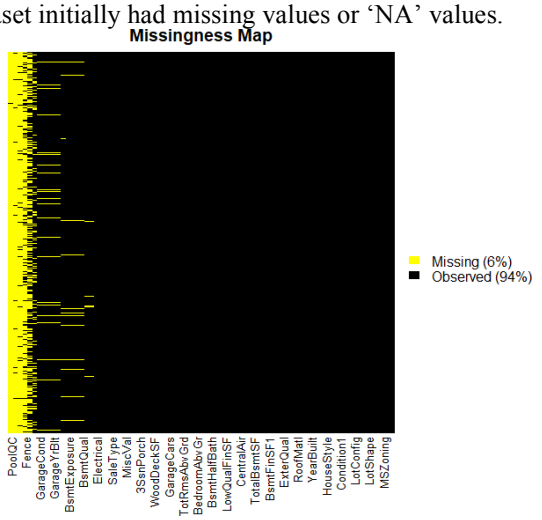
ThedatasetusedinthisprojectcomesfromtheIowahousing data taken from Kaggle. The dataset is in csv(comma separated values). It contains about 1460 entriesrepresenting the aggregate information about 80 features of homes from various suburbs located inIowa,USA.

Some of the features are:

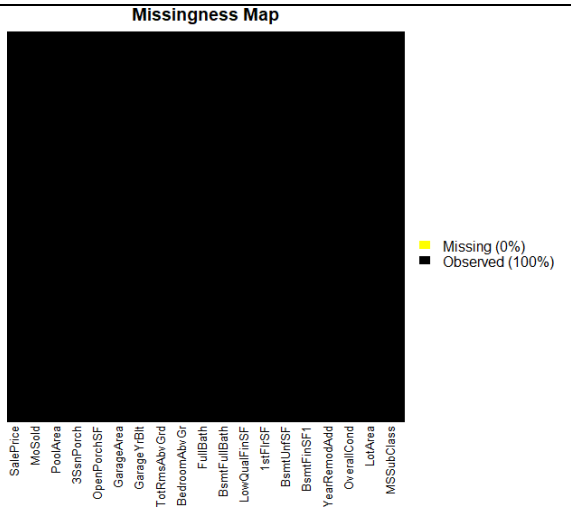
1. SalePrice - the property's sale price in dollars. This would be taken as the target variable to bepredicted.
2. OverallQual: Overall material and finishquality
3. GrLivArea: Above grade (ground) living area squarefeet
4. YearBuilt: Original constructiondate
5. GarageArea: Size of garage in squarefeet
6. FullBath: Full bathrooms abovegrade
7. TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)
8. BedroomAbvGr: Bedrooms above grade (does NOT include basementbedrooms)
9. PoolArea: Pool area in squarefeet

B. DATACLEANING:

Data cleaning is the process of preparation of data for analysis by modifying or removing data that is incomplete, incorrect, duplicated, improperly formatted or irrelevant and then re-modelling, or removing the dirty or crude data.
The dataset initially had missing values or 'NA' values.



na.omit(data) function removes all those rows which have NA value.

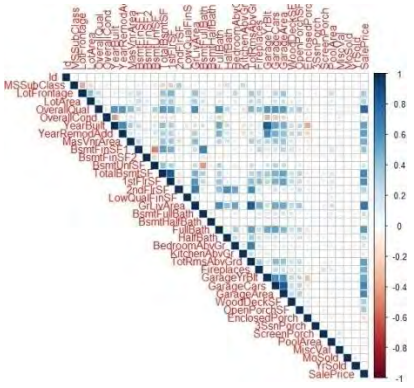


It returns a data table with just the rows where the specified columns have no missing value in any of them.

C. ANALYSINGDATA:

In order to see which columns are to be used for model fitting for Linear Regression and Decision Tree, we use a library called as 'corrplot'. It has a function called as corrplot which returns a graphical plot demonstrating the correlation between different variables of the dataset.

The variables showing a high positive correlation with the dependent variable are extracted for model fitting.



It is observed that the following variables showed the highest correlation:

- GarageArea
- GarageCars
- TotRmsAbvGrd
- FullBath
- GrLivArea
- TotalBsmtSF
- OverallQual
- YearBuilt
- Fireplaces

As a result, we use the above variables for our Linear Regression and Decision Tree model fitting.

SalePrice	OverallQual	YearBuilt	GrLivArea	TotRmsAbvGrd	GarageArea
208500	7	2003	1710	8	548
181500	6	1976	1262	6	460
223500	7	2001	1786	6	608
140000	7	1915	1717	7	642

D. APPLYINGMODEL:

Two machine learning models were used: Decision Tree and Multivariate Linear Regression

Multivariate **Linear Regression**:

In order to implement Multivariate Linear Regression, the lm() function is used.

Syntax of lm() function:

lm(formula, data, subset, weights, na.action, method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE,
singular.ok = TRUE, contrasts = NULL,offset, …)

formula: It is a symbolic description of the model to be fitted.

data: This is the dataset being considered.

model, x, y, qr -the corresponding components of the fit are returned if TRUE

DECISION TREE:

In order to implement Decision Tree the library rpart is used.

rpart function is used which used to plot a decision tree.

rpart(formula, data, weights, subset, na.action = na.rpart, method, model = FALSE, x = FALSE, y = TRUE, parms,control, cost, …)

E. PERFORMANCE MEASURES OF THEMODELS:

MULTIVARIATE LINEAR REGRESSION:

The R-Squared value, also called as the coefficient of determination, is checked in linear regression which represents the proportion of the variance for a dependent variable (House Price) that's explained by independent variables.

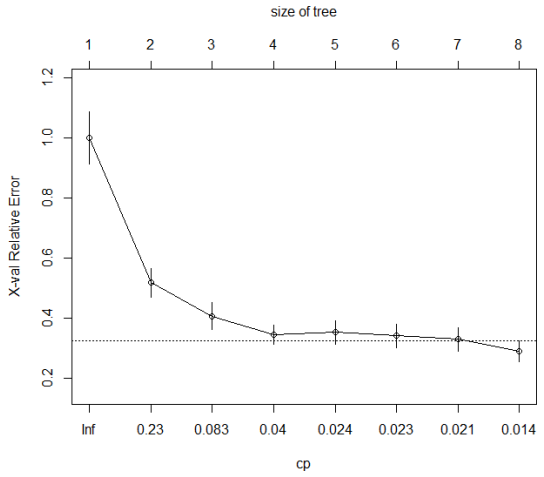
R-squared value measures how good our model has been fitted to the given training data. It evaluates the scattered data points around the fitted regression line. It is defined as the ratio of the variance explained by the model to the total variance. Higher R-squared values indicate that the difference between observed data and fitted values is small. R-squared is always measured between 0 and 100%.

The R-squared value of the linear regression model can be determined by using the summary(model) function which gives the residual error along with the R-squared value.

DECISION TREE:

For decision tree we use a plot function given as plotcp(model) which draws a line representing the highest cross validated error less than the minimum cross validated error plus the standard deviation of the tree.

Cross validation creates a number of partitions from the sample observations. These are called as validation sets and are obtained from the training data set. After fitting the model on training data, the performance is measured against each validation set and averaged. This average value gives a better assessment of how the model will perform to predict for observations



III. LITERATURESURVEY

A. HOUSE PRICE AFFECTINGFACTORS:

In the research paper that we referred [2] , R.A. Rahadi broadly classifies the house price affecting factors into 3 categories: physical conditions, concept andlocation.
Physical conditions include the properties of a house that can be perceived by human senses such as size, number of bedrooms, kitchen and garage availability, garden and area of land, age of the house [3] etc. Concept is a design scheme used by developers to attract buyers. These could include healthy and green environment, minimalist homeetc.
Location is one of the most important factors that affects the price of a house. The price of the house changes with respect to its proximity to public facilities such as schools, hospitals, malls, health centers, parksetc.

B. HEDONIC PRICING:

This is a price prediction model based on the hedonic price theory. The hedonic price theory assumes that the value of a property is the sum of all its attributes value [4]. For the implementation, hedonic pricing can be implemented using regression model. y = a.x1 + b.x2 + c.x3 +... + n.xi Where, y is the predicted price, and x1, x2, … xi are the attributes of a house. a, b, ... n are the correlation coefficients associated with each variable respectively for the determination of house prices. Sirmans, Macpherson and Zietz (2005) provide a study of 125 papers that use hedonic pricing model to estimate house prices in the past decade [5]. The paper gives a list of 20 attributes frequently used to specify hedonic pricing models. Our dataset contains 80 variables. The paper also discusses the effects of variables on the house prices. For example, number of bath-rooms is usually positively correlated to the final sale price. In our model we found that some variables such as garage area, pool size, the year in which the house was built, presence of a fireplace, the overall quality etc. had a very strong positive correlation with the sale price of thehouse.

IV. FLOWCHART

V. RESULT ANDDISCUSSION:

Accuracy:

TABLE 1

Expected SalePrice	Predicted SalePrice	Error Range
208500	255249.1	20-25%
181500	111691.8	
223500	344341.3	
140000	103489	
250000	239014.8	

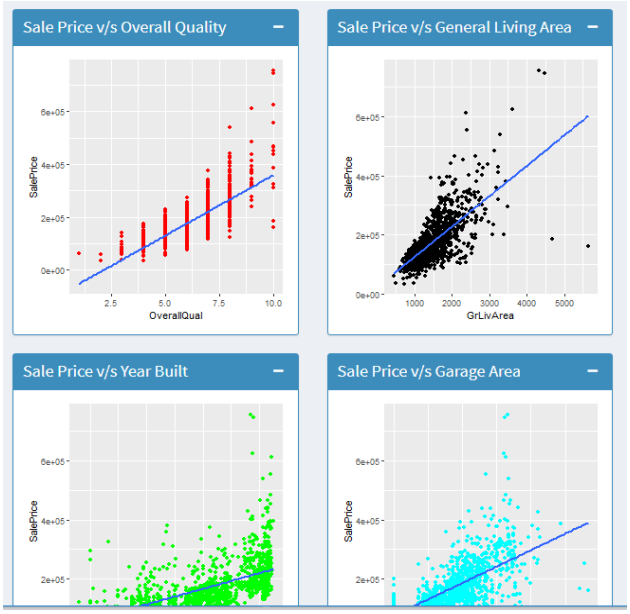
The error obtained when the values are predicted by multivariate linear regression is 20-25% (Table 1)

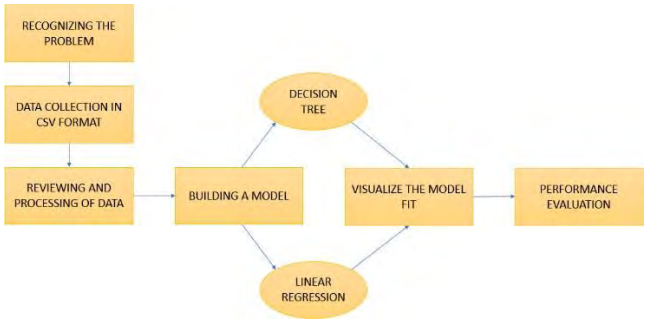
TABLE 2

Expected SalePrice	Predicted SalePrice	Error Range
208500	261186.8	15-20%
181500	125166.3	
223500	310979.5	
140000	189614.3	
250000	189614.3	

The error obtained when the values are predicted by decision tree is 15-20% (Table 2)

We implemented shiny web app in our project in order to make it a more user-friendly append we got the following results :





VI. CONCLUSION:

Real Estate is considered one of the most lucrative areas of investment. It can however be overwhelming for a beginner to understand the properties and features of a house which contribute to its value. Our goal was to create something that while being user friendly provides the operator with the necessary information in a clear concise manner to make an informed choice.

VII. References

[1] (IJACSA)InternationalJournalofAdvancedComputerSciencea nd Applications,
Vol. 8, No. 10, 2017 :

[2] R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartotoor, and I. B. Syamwil, ―Factors influencing the price of housing in Indonesia,‖ Int. J. Hous. Mark. Anal., vol. 8, no. 2, pp. 169–188,2015.

[3] V. Limsombunchai, ―House price prediction: Hedonic price model vs. artificial neural network,‖ Am. J. …,2004.

[4] S. Rosen, ―Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,‖ J. Polit. Econ., vol. 82, no. 1, pp. 34– 55,1974.

[5] G. Stacy Sirmans, David A. Macpherson, and Emily N. Zietz. "The Composition of Hedonic Pricing Models". In:Journal of Real Estate Literature13.1 (2005)

A Review on Dynamic Airline TicketPrice Prediction using Machine Learning

Anurag Tiwari,Janhavi Tawde
Department of Computer Engineering
Thakur College of Engineering and Technology, mumbai
anurag219tiwari@gmail.com,tawwjazz267@gmail.com

Abstract—— **this paper aims to review the dynamic pricingof anairlineticketforaflightonvariousroutes.Itproposesadata-drivenmodeltopredictthebesttimeforacustomertopurchase the tickets with the use of Machine Learning. The changing airfare is dependent upon various factors which are critically analyzed. This paper aims to study the factors which influence the fluctuations in the airfare prices and predict the prices of individualflights.**

Keywords—dynamic airfare, price prediction, machine learning

I. I. INTRODUCTION

Today, the airline corporations are using complex strategies and methods to assign airfare prices in a dynamic fashion. These strategies are taking into account several financial, marketing, commercial and social factors closely connected with the final airfare prices [3]. Any individual who has booked a flight ticket previously knows how unexpectedly and invariably the cost of a ticket can change. This is due to the fact that the airline companies use convoluted methods and mathematical models to assign the charges in a dynamic fashion. They use advanced strategies called Revenue Management or "yield management" to execute a distinctive valuing strategy [1]. The least expensive ticket experiences change in its value over a period of time. Conventional variables such as distance, fuel prices although play an important role, are no longer the only exclusive factors that dictate the pricing strategy. Elements related to economic, marketing and societal trends have played increasing rolesin imposing the dynamic prices. Due to the high complexity of the pricing models applied by the airlines, it is very difficult for a customer to purchase an air ticket at the lowestprice.

The Revenue management strategy aims to maximize the profit of the flight companies. It mainly works on time of purchase patterns and having the plane occupied as per their need. The time of purchase patterns indicate that the last- minute purchases will be expensive. The companies can dictate the occupancy of the seats in a flight by raising the cost on a flight which is filling fast in order to reduce sales and so, keeping some seats unoccupied for those last-minute expensive purchases. The intensity of optimizing the profits is huge which leads a consumer in paying more for a similar seat. According to a report, India's civil aviation industry is on a high-growth trajectory.

India aims to become the third- largest aviation market by 2020 and the largest by 2030. Indian domestic air traffic is expected to cross 100 million passengersbyFY2017,comparedto81millionpassengersin

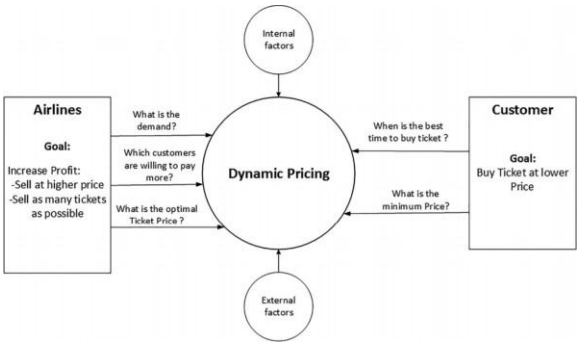
A. Selecting a Template (Heading2)

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file.2015,asperCentreforAsiaPacificAviation(CAPA).Asth e workingclassofthecountryisexposedtoairtravelthesearch for the least expensive air ticket becomes inevitable.[1]Due to the high complexity of the pricing models applied by the airlines, it is very difficult a customer to purchase an air ticket in the lowest price, since the price changes dynamically.Forthisreason,severaltechniques[3],[4],able toprovidetherighttimetothebuyertopurchaseanairticket by predicting the airfare price, have been proposed recently. The majority of these methods are making use of sophisticated prediction models from the computational intelligenceresearchfieldknownasMachineLearning(ML). [3]Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) make it possible to infer rules and model variations on airfare price based on a large number of features, often uncovering hidden relationships amongst the features automatically. To the best of our knowledge, all existing work leveraging machine learning approaches for airfare price prediction are based on: 1) proprietary datasets that are not publicly available [8] [9] and 2) transaction records data crawled from online travel booking sites like Kayak.com [10] [11] [12]. [4]

The issue with the latter is that the transaction records from eachonlinebookingsiteareasmallfractionofthetotalticket sales from the entire market, making the acquired data likely tobeskewed,andthus,notrepresentingthetruenatureofthe entire market.[4]

Fig 1 – Dynamic Pricing

II. LITERATURESURVEY



There have been several studies conducted both at the airline and customer side but there has been no effort made to presentanoverviewoftheexistingwork.Theref ore,themain goal of this paper is to present a comprehensive literature review of existing

studies related to this topic which can be utilized by future researchers. Several issues havebeen

Discussed including data sources, features and various techniques employed for prediction. We believe that this is an important contribution for researchers who are aiming to work on this exciting area of research. [2]. One of the results ofourreviewindicatesthatexistingmodelsgenerall yrelyon limited number of features which are not effective enough in predicting ticket price. For example, customer side models generally utilize restricted features extracted from historical ticketpricedata,ticketpurchasedateanddepartured ate.Ina similarway,airlinessidemodelsarealsodevelopedb asedon limited internal factors such as seasonality, holidays, supply (number of available airlines and flights), fare class, availability of seats, recent market demand, flight distance and competitive moves by other airlines etc.[2]

Nowadays, social media sentiment analysis has become a good source of information. For example, social media data has been used for event prediction competitor intelligence, price prediction and tourist traffic flow prediction. A similar approach could be followed to extract useful social media information related to various external factors affecting airlines passenger demand and ticket price. For example, analysis of different twitter hash tags could give valuable information about the presence of an event at an origin/destination city, competitors' promotions, volume of tourist traffic flow, weather condition, economic activity etc. This in turns might allow us to predict the change in ticket price/demand. It is expected that a data mining model that utilizes information resulting from social media data would give better results than existing work in forecasting route demand and or ticket price. However, to the best of our knowledge, there is no existing work that utilizes social media data to predict route demand and or ticket price.

In this paper we have analyzed three different approaches [1][3][4]tothesameproblemofairlineticketpricepr ediction and have reviewed a fourth one [2]..

III. COLLECTINGDATA

There are hardly adequate research papers available discussing the various ticket pricing strategies that are implemented by several airlines and Online Travel agencies (OTA. This can be due to two reasons: first, ticket pricing strategies are highly business sensitive and remain proprietary of the owner company [2]. Most airlines prefer not to reveal their ticket pricing strategies because of competition with other airlines. Second, there is lack of publicly available datasets that could enable researchers to conduct their prediction effectively. As a result, researchers are obliged to rely on small datasets that are sometimes gathered using web scrapping programs. [2]

The three different approaches we studied have been classified as model 1, 2 and 3 and their data collection strategies are discussed.

1. Model 1

There were various sources of the data on the web which can be utilized for training the models. Websitesgiveinformationaboutthemultipleroutes,

Times, airlines and fare. To implement this data collection, a website "makemytrip.com" was selected and python is used for the implementation of the models and collection of the data [5].

a) Collection ofdata

The script extracts the information from thewebsite and creates a CSV file as output. This file contains the information with features and its details [13]. The following features are considered for the scarping ofdata

- Origin
- Destination
- DepartureDate
- DepartureTime
- ArrivalTime
- TotalFare
- Airways
- Taken Date

In this study, the focus was only on minimizing the airfare charges. Therefore, a single route was considered without return. This data is collected for one of the busiest routes in India (BOM to DEL) over a period of three months that is from February to April. For each flight data with all the features was collected manually.

b) Cleaning and preparing data

The collected data was preprocessed and cleaned in order to remove duplicate and null values from the datasets.

c) Analyzing data

Data Preparation was followed by analyzing the data,uncoveringthehiddentrendsandthenapplying various machine learning models. Days todeparture was calculated from the day the data was taken and the departure date. This was considered to bewithin therangeof45days.Thedayofdeparturewasfound to be a significant feature affecting the price. For example, the price of the airline ticket was cheaper onaweekdaythanaweekend.So,thetimehadbeen divided into four categories: morning, afternoon, evening,night.

2. Model2

Initially, the Greek Aegean Airlines [7] company and its flight, from Thessaloniki to Stuttgart, was selectedasthecasestudyoftheinvestigationforthis model. The current study consisted of four distinctivephases:

- the selection of the features of the flight that influence the airfareprices
- the collection of enough flights data which willbeusedtotrainandtesttheappliedML models,
- selection of the regression ML models being

compared
- ExperimentalevaluationoftheMLmodels.

During the phase of feature selection, the most informative features of a flight that determine the prices of the air tickets are decided. This phase was an extremely crucial phase as it defines theproblem under solving. For every flight the following features wereconsidered:

- F1: Feature 1 - departure time.
- F2: Feature 2 - arrivaltime.
- F3: Feature 3 - number of free luggage (0, 1 or 2).
- F4: Feature 4 - days left untildeparture.
- F5: Feature 5 - number of intermediatestops.
- F6: Feature 6 - holiday day (yes orno).
- F7: Feature 7 - overnight flight (yes orno).
- F8: Feature 8 - day ofweek.

The feature F4 indicates the number of days between the ticket purchase and the day of the flight.

In this study, the interest was focused on the prediction of a single airfare price without return. Forthesakeoftheexperimentsasetofflightstothe samedestination(fromThessalonikitoStuttgart)for the period between December and July, was collected. For each flight the eight features (F1:F8) were manually collected from the Web, 1814flights were recorded totally.

2. Model 3

In order to build the airline ticket price model at the market segment level, information about both the airline traffic and passenger volume for eachmarket segment is required. Therefore, two public datasets (DB1B and T-100) are used in our proposed framework. Data collected during 2018 are used to train and evaluate the proposed model. Table I summarizes the information of these two datasets. The United States Department of Transportation regularlyupdatesboththeDB1BandtheT-100.The DB1B dataset provides quarterly-aggregated information about the airline tickets in the United States from reporting carriers and consists of 10% randomly sampled ticket data from each reporting carrier. The information in DB1B is organized in three parts, namely “Coupon”, “Market”, and “Ticket. For our proposed framework, a subset of most related data are used, including the origin airport (ORIGIN), the destination airport (DEST), time of the itinerary (QUARTER), carrier information (REPORTING CARRIER), seat class (SEAT CLASS) (e.g., first, business, economic, etc.), total flight distance for a ticket (DISTANCE), airfare price (ITIN FARE), and the number of passengers in a ticket (PAX). Different fromDB1B, T-100 provides monthly domestic non-stopsegment data reported by both the domestic andinternational carriers.Itpresentsthenumberofpassengersofeach airline and each market segment by aircrafttype.

The selected features with their description which were chosen for this study are displayed in Table 2.

Entity	Availability	Data
Ticket	DB1B	fare price, total distance, and total number of passengers
Coupon	DB1B	market segment, time of theitinerary, carrier, and seatclass
Market Segment	DB1B&T-100	original airport, destination airport, and segment distance
Market Segment by carrier	T-100	number of passengers, and number of available seats by aircrafttype

Table 1 – Data entities and values

Feature Name	**Description**
Distance	Market distance between the origin and destination airports
Seat Class	Indicator for economy or premium seat type
Passenger Volume	Total number of passengers traveled between the origin and destination airports
Load Factor	The ratio of the total number of passengers to the total number of seats in a market
Competition Factor	The market HHI
LCC Presence	Indicator of LCC operating in the market
Crude Oil Price	Quarterly average of crude oil price
CPI	Quarterly average of Consumer Price Index
Quarter	Indicates the three-month period of the year

Table 2 – Feature Selection

III. MACHINE LEARNINGMODELS

1. Model 1

To develop the model for the flight pricepredictionmany conventional machine learning algorithms are evaluated. They are as follows: Linear regression, Decision tree [8], Random Forest Algorithm [9], K- Nearest neighbors [11], Multilayer Perceptron [10], Support Vector Machine (SVM) [12] and Gradient Boosting. All these models are implemented in the scikit learn. To evaluate the performance of this model, certain parameters are considered. They are as follows: R-squared value, Mean Absolute Error (MAE) and Mean Squared Error (MSE). The formulasforthesethreeparametersareasfollows:

$$R^2 = 1 - \frac{\sum_n^{t=1}(yi - \hat{yi})^2}{\sum_n^{t=1}(yi - \bar{yi})^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{n}^{t=1}|yi - \hat{yi}| \quad (2)$$

$$MSE = \frac{1}{n}\sum_{n}^{t=1}(yi - \hat{yi})^2 \quad (3)$$

A. LinearRegression

Regression is a method of modeling a target valuebasedonpredictorsthatareindependent. Itismostlybasedonthenumberofindependent variables and the relationship between independent and dependent variables. Linear regression is a type of analysis where the numberofindependentvariablesisoneandthe relationship between the dependent and independent variables varylinearly.

$$y\ (pred) = b0 + b1 * x \quad (4)$$

A. Decision tree

The Decision tree calculation separates the informational collection into smallsubsets, at a similar same time it creates gradually. The entire informational collection is considered as the root initially. The characteristic qualities determine their dispersion. There are two primary characteristics in the decision tree calculation: Information Gain and Gina index. Information Gain is the proportion of change in entropy. Gina Index is a component that measures how frequently an arbitrarily picked component would be mistakenly distinguished. It implies a characteristic with a lower Gini index ought to beliked.

B. RandomForest

It is a supervised learning algorithm. The benefit of the random forest is, it very well may be utilized for both characterization and relapse issue which structure most of Current machine learning framework. Random forest forms numerous decision trees, what's more, adds them together to get an increasingly exact and stable expectationItisverysimpletodiscoverthe significance of each element on the expectation when contrasted with others in this calculation. [9]

C. K-NearestNeighbors

In regression techniques, the output obtainedisanaveragevalueofitsknearest neighbors. It is a non-parametric method like SVM. Using some values, the results are evaluated and the best performance value isobtained.

D. Multilayer perceptron(MLP)

It is the class of feed forward artificial neuralnetworks.Itincludestheinputlayer, output layer and the number of the hidden layers. The hidden layer gives the depth of the neural network. The setup includes 1 hidden layer, the number of neurons starts from 100 to 2000 with different intervals depending upon the required condition. To fire each neuron, it requires activation energy. The logistic sigmoid function is used as an activationfunction.

E. Gradientboosting

Itisanadditiveregressionmodelbyfitting simple function to current "pseudo" residuals sequentially by least-squares at eachiteration.ItusestheDecisiontreeasa basic estimator in scikit implementation. Starting from 10 to 1000 with the interval of 10 boosting stages are used with maximumnumbers

Thelossfunctionisanimportantparameter in the gradient boosting. It can be calculated with options: least squares regression, least absolute deviation, and quartile regression. The most important featureintheflightpricingpredictionisthe day is a holiday or not, the day is weekday or weekend and the difference between the days.

D. Support Vector Machine (SVM)

In the paper Support VectorMachine used as regression analysis that relays on kernel function considered as non-parametric technique [2]. The following kernels are used: Linear, Polynomial, Radial Basis Function [12]. As per the previous studies Random forest and the gradient boosting gives the maximum accuracy[11].

2. Model 2

Eight state of the art regression ML models [10], [12], [13], [14], [15] were selected for the current study and applied to the same data of flights. The MLmodelscomparedinthisworkarethefollowing:

- Multilayer Perceptron(MLP).
- Generalized Regression NeuralNetwork.
- Extreme Learning Machine(ELM).
- Random Forest RegressionTree.
- Regression Tree.
- Bagging RegressionTree.
- Regression SVM (Polynomial andLinear).
- Linear Regression (LR)

The 1814 flights collected in phase 2, were used in a 10-fold cross-validation procedure to train the aforementioned ML models. The performance indices used to compare the models are the prediction accuracy (% - MSE between the desired and predicted prices) and the time in seconds, needed to train each model.

IL Model	**onfiguration**
Multilayer Perceptron (MLP)	hidden layers 5 nodes each layer
Generalized Regression Neural Network	pread=1.0
Extreme Learning Machine	0 neurons
Random Forest Regression Tree	00 weak classifiers (decision trees)
Regression Tree	linParentSize=10 MinLeafSize=3 MaxNumSplits=45
Bagging Regression Tree	00 weak classifiers (decision trees)
Regression SVM (Polynomial)	rder=3
Regression SVM (Linear)	ochastic gradient descentsolver
Linear Regression	ual stochastic gradient descent solver

Table 3 – Models Configuration

3. Model 3

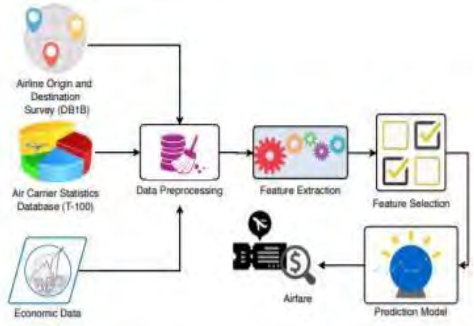


Fig 2 – Proposed Framework

For the experiments, samples of 16,577,497and 41,360,566 were collected from the 2018DB1B ticket table and coupon table, respectively. The T-100 dataset contains 329,426 samples. Several well-known machine learning models were tested as baselines to compare with theRF Model. In particular, LR, SVM, Multilayer Perceptron (MLPs), and XGBoost Tree were used for the evaluation.

For the SVM model, the radial basis function kernel is used, the tolerance for stopping criterion is set to 0.001, and the penalty parameter for the error term is set to 0.1. For the MLPs, three hidden layers are used with 30 neurons per layer. The Rectified Linear Unit (ReLU) is used as the activation function and Adam is the optimization function. The learning rate is set to 0.0001 with momentum enabled set to 0.9. For the XGBoost model, the number of estimators is set to 100 with a learning rate as 0.1, and max depth equals to 5. For the RF model, the number of estimators is also set to 100 with the minimum number of samples to split set to 2. To evaluate the proposed price prediction model, two popular metrics for regression analysis are used: the Root Mean Square Error (RMSE) and the Adjusted R Squared. RMSE calculates the differences between the observed values, y, and predicted values, y^. This difference for each data point is also called the residual. Thus, RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

Where N is the total sample size. The lower the RMSE value is, the higher performance the regression model has. The Adjusted R Squared, (R2 adj), is usually used to explain how well the independent variables fit a curve or line. Adjusted R2 also adjusts for the number of variables in a model. The higher the Adjusted R Squared is, the better the result of regression is. It is calculated as follows:

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(N-1)}{N-p-1}\right]$$

where p is the number of predictors and R2 is:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

Here $\bar{y}$ is the mean value of y.

ML algorithms	R-squared	MAE	MSE
Random forest	0.67	0.08	0.04
Multilayer Perceptron	0.65	0.09	0.04
Gradient Boosting	0.47	0.13	0.06
Decision tree	0.45	0.09	0.06
K-nearest neighbour	0.38	0.14	0.07
SVM	0.19	0.15	0.08
AdaBoost	-0.12	0.21	0.11

Table 4 – Algorithm Evaluation

After evaluating the performance of the all machine learning models, further improvements were made usingacorrectpredictormodelforthebestresult.Two separated train models were developed by applying the trained datasets. Also the appropriate weights are assigned to them to get a better predictor model[16].

A. Stacked PredictionMode

The performance of machine learning models were evaluated critically. The study discovered that Random forest and Multilayer Perceptron, have better results compared to other models. Some weights are applied on the prediction results of these two models to get better prediction results.
This was a stacked predictionmodel.

B. Exhausted searchmethod:

This method found the optimal values for theta by running an exhausting search over a domain of finite interval from -n ton at the difference of a particular stepsize.

2. Model2

The influence of some critical features will be examined through a "one leave-out" rule. A 10- fold cross-validation procedure was applied to all the experiments and the mean performance of each model is presented in this section. The performance of all models for the case of the entire feature set (eight features) is presented in the table below, with the highest performed model being bold faced.

ML Model	**Accuracy(%)**	**Execution Time (sec)**
Multilayer Perceptron	80.28	20.88
Generalized Regression Neural Netwo	66.83	0.13
Extreme Learning Machine	68.68	0.05
Random For Regression T	85.91	5.50
Regression Tree	84.13	0.04
Bagging Regression Tree	**87.42**	17.05
Regression SVM (Polynomial)	77.00	1.23
Regression SVM (Linear)	49.40	0.34
Linear Regression	57.25	0.10

Table 5 – Results with all features

The following table highlights the changes inmodel when one leave-out rule was applied and some critical features were deliberately eliminated. The conclusive outcome on the model was as shown in thetable.

Feature Eliminated	Changes in Model	Conclusion
Departure and Arrival Time	Almost all models shown lower (up to 10%) prediction accuracy and greater execution time.	"departure time" and "arrival time" influence significantlythe Airfare prices.
"day of week"	all models were not affected as much as previously, except "Regression SVM" With Linear kernel.	"day of week" does not influence airfare prices
"overnight flight	The models perform similarly or even worse with the case of using all features. Only the "Multilayer Perceptron" and the "Regression SVM" with Linear Kernel seems to be affected significantly bythis Feature.	this feature is not related with the price of the air ticket,
holiday day	The "Bagging Regression Tree" outperforms all the models not only inthis experiment, but also all the modelsunder Different feature sets examinedpreviously.	The reminder models seem Not to be affected by the exclusion of "holiday day" feature.
number of intermediate stops	almost all models shown lower (up to 10%) prediction Accuracy and greater execution time.	Number of intermediate stopsinfluence Significantly the airfare prices.

Table 6 - Eliminated Features Conclusions

Concluding the previous study, one can claim that "Bagging Regression Tree", "Random Forest Regression Tree", "Regression Tree" and MLP modelsarethemoststablemodelsaccordingtotheir accuracy scores. In addition, as far as the execution time is concerned the best models are "Random Forest Regression Tree" and "Regressiontree".

3. Model 3

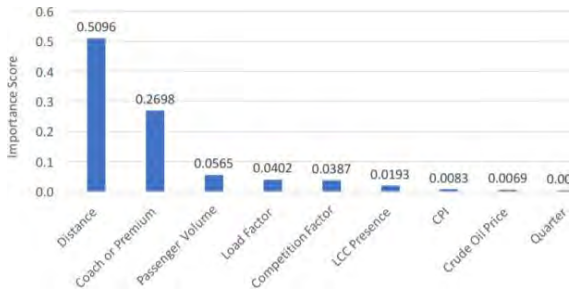


Figure 3_ - Importance of Feature generated by RF

In order to demonstrate the importance of each feature for airfare price prediction, the importance scores generated by the feature selection module were retrieved. The figure depicts the importance value for each feature. As shown in the figure, "Distance" and "Seat Class" (Economy or business) are the most important factors for airfare price estimation followed by "Passenger Volume", "Load factor", and "Competition Factor". Although the "CPI" and "Crude Oil Price" do not have very high importance scores, they can still help the model predict a more accurate estimation of the airfare price. However, based on the experiments, "Quarter" does not help the regressionmodel. Including the variable "Quarter" does not reduce the error during the training phase. Thus, it is automatically removed by the RF feature selection module.

The results comparing various regression models with feature selection and without feature selection areshowninthetable.Ascanbeseenfromthistable, LRandSVMhavethelowestperformancecompared to other ML methods with respect to the RMSE and R^2adjmetrics.Theperformanceofallofthemodels improves after applying feature selection, which illustrates the importance of this module. XGBoost performs better than MLP, SVM, and LR, but does not outperform RF for airfare price prediction. Therefore,weutilizeRFintheproposedframework, whichachievesthehighestperformancecomparedto other baselines for this dataset. Specifically, it reaches 62.753 and 0.869 RMSE and R^2 adj , respectively. In other words, it improves the R^2 adj by 40% compared to the LR model, which is extensively used in the previous studies for airfare pricepredictionAlthoughthelessimportantfactors may not contribute significantly to theperformance, theseresultsshowthattoachievethebestperforming model, one should include the "Load factor", "CompetitionFactor","CPI",and"CrudeOilPrice" as features. Consequently, the proposed framework utilizes all of these features to achieve the highest airfare price predictionperformance

Method	Without feature selection		With feature selection	
	RMSE	R^2 adj	RMSE	R^2 adj
LR	111.000	0.612	110.284	0.618
SVM	112.963	0.587	108.358	0.626
MLP	88.447	0.754	85.832	0.766
XGBoost	83.481	0.778	80.447	0.797
RF	66.584	0.858	62.753	0.869

Table 7 - Performance Comparison for different Regression Models with and without feature selection

V. CONCLUSION

This paper reported on a preliminary study in "airfare prices prediction". In the paper the overall survey for the dynamic price changes in the flight tickets is presented. The machine learning models which were reviewed in this paper utilized all different machine learning algorithms to be applied on different datasets as and where they were collected from.We observed that as the pricing models of the company are developedinordertomaximizetherevenuemanagement,the various features that influence the prices of the ticket are to beconsidered.Weobservedthatitisfeasibletopredictprices for flights based on historical data using machine learning. The experimental results show that ML models are a satisfactorytoolforpredictingairfareprices.Otherimportant factorsinairfarepredictionarethedatacollectionandfeature selectionfromwhichwedrewsomeusefulconclusions .From the experiments we concluded which features influence the airfare prediction at most. Apart from the features selected, there are other features that could improve the prediction accuracy.

VI. REFERENCES

[1] Supriya Rajankar, Neha Sakharkar, "A Survey on Flight Pricing Prediction using Machine Learning" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 8 Issue 06,June-2019

[2] Juhar Ahmed Abdella, Nazar Zaki, Khaled Shuaib, Fahad Khan, "Airline Ticket price and demand prediciton: A survey", Journal of King Saud University – Computer and InformationSciences

[3] K. Tziridis, Th. Kalampokas, G.A. Papakostas, K.I. Diamantaras, "Airfare Prices Prediction Using Machine Learning Techniques", EuropeanSignalProcessingConference(EUSIPCO),28August 28-2 September, Kos, Greece,2017.

[4] TianyiWang∗,SamiraPouyanfar∗,HaimanTian∗,YudongTao† ,"A Framework for Airfare Price Prediction: A Machine Learning Approach"

[5] www.makemytrip.com

[6] K. Tziridis, Th. Kalampokas, G.A. Papakotas and K.I. Diamantaras,"Airfare Prices Prediction Using Machine Learning Techniques", EUSIPCO2017.

[7] Aegean Airlines,https://en.aegeanair.com

[8] S.B. Kotsiantis, "Decision trees: a recent overview," Artificial Intelligence Review, vol. 39, no. 4, pp. 261-

283,2013

[9] L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5-32 , 2001.

[10] S. Haykin, Neural Networks – A Comprehensive Foundation. Prentice Hall, 2nd Edition, 1999.

[11] Viet Hoang Vu, Quang Tran Minh and Phu H. Phung,”An Airfare Prediction Model for Developing Markets”, IEEE paper 2018

[12]

[13] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola and V. Vapnik, ”Support vector regression machines,” Advances in neural information processing systems, vol. 9, pp. 155-161, 1997.

[14] G.B. Huang, Q.Y. Zhu and C.K. Siew, “Extreme learning machine: Theory and applications,” Neurocomputing, vol. 70, no. 1-3, pp. 489- 501, 2009.

[15] G.A. Papakostas, K.I. Diamantaras and T. Papadimitriou, “Parallel pattern classification utilizing GPU-Based kernelizedslackmin Algorithm,” Journal of Parallel and Distributed Computing, vol. 99, pp. 90-99, 2017.

[16] L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and Regression Trees. Boca Raton, FL: CRC Press,1984.

[17] Yudong Tan and Huimin Zhou,”A Bayesian Predictor of Airline Class Seats Based on Multinomial Event Model”,IEEE on Big Data2016

Exploiting Processor Variability for Ecient Application Mapping

Akash Yadav[1], Mushtaq Ahmed[1], and Daitule Omkar Vilas[1]
Malaviya National Institute of Technology, Jaipur, Rajasthan, India
2019rcp9150@mnit.ac.in mahmed.cse@mnit.ac.in

Abstract. **Increasing on-chip integration with technology scaling in which more transistors are available on a chip than that can be simultaneously powered on will lead to the dark silicon era. Process variation gradu-ally impacts the performance (power, speed, frequency etc.) of system over its usage life. In this paper, by adapting the application allocation algorithm at the software layer, reduction of the impact of process varia-tion in multicore architecture is suggested and tested. Process variation causes core-to-core variations in power and frequency. To gain the maxi-mum performance within the power limits, best subset of cores should be selected from the available cores to run an application. When the power constraints are used as cost parameter, simulation of adaptive alloca-tion algorithm on sniper multicore simulator promises twice the better performance compared to conventional core allocation strategy.**

Keywords: Process Variations Dark Silicon E cient Application Mapping Sniper multicore simulator.

I. INTRODUCTION

Process variation [1] is naturally occurring variation in attributes of transistors, which results in measurable variance in the output performance and behavior of all circuit. We tried to execute processes in a way that it will choose cores based on its timing and maximum power constraints, which will be adaptive to the changes in the features of core like frequency. The basic idea here is to optimize the execution in such a way that the ine cient cores in the pool are turned o instead of the ones which are fast and which can give us better performance. This process of picking better cores out of available cores is called cherry picking [2].The focus of this paper is on exploiting process variations in dark silicon cores for performance improvement under a power budget so as to solve the problem of variation-aware core selection.

II. LITERATURE SURVEY

With the advancement of the technology, transistors are becoming faster and smaller, but transistor power consumption does not scales with the same level [3]. Due to supply voltage limits, power densities rapidly increase on the chip. So to prevent thermal emergencies ,a signi cant amount of on-chip resources needs to stay dark, This phenomenon is known in the literature as dark silicon [4, 5]

.Process variation and dark silicon are the most important factors impacting the performance of multicore systems in today's world.

2.1 Process variation

Mostly post-manufacturing tuning is used to reduce the impact of process vari-ation [6]. This conventional method often leads to system overdesign relative to their speci cation. Less leaky processing units are more energy e cient than their leakier counterparts at a given supply voltage and frequency. Algorithm suggested in this paper takes advantage of this observation, to shift the pro-cesses to execute on less leaky ones, maintaining performance while reducing total power consumption.

2.2 Cherry-picking: process variation aware core selection

Increased number of cores on a single chip led to dark silicon era, it refers to the amount of silicon that cannot be powered-on at the nominal operating volt-age for a given thermal design power (TDP) constraint. Certain cores of the system remain unpowered to avoid overheating. According to recent studies at 8 nm technology nodes, the amount of Dark Silicon may reach up to 50%-80% depending upon the processor architecture, cooling technology and application workloads. Some early work has been done in this eld which uses a heteroge-neous approach of using Hardware accelerators to get better performance out of the system [7]. When accounted for Process Variations, the approach becomes quite di cult.

III. PROPOSED WORK

An adaptive application allocation algorithm is proposed which can be used at any level of the core to application allocation. It uses two algorithms, one for estimation of execution cost and another for allocation of a particular core to a particular application so as to minimize the total cost of running these applica-tions. Input to these algorithms are application statistics, core frequencies and cost of running each application on di erent available cores. It rst gets the appli-cation statistics like maximum runtime allowed for each application to maintain the quality and minimum frequency requirements for each application. Minimum required frequency is stored as f_{min} for all i applications. It then estimates cost (here taken from sniper- runtime dynamics values) of using a particular core for a particular application based on estimation algorithm. Core frequencies are read and stored in f_{core} for all the j cores. It then checks whether a particular core can be used for a particular application by checking it against maximum runtime requirements. If it doesnt satisfy the constraints of the application then its cost is multiplied with a large value so that this particular core can never be used for this application. After all applications have been checked against all cores, it allocates a particular core to a particular application by using an allocation algorithm (Munkres allocation algorithm is used in this paper)[8]. It then displays the optimal mapping of each core with a certain application so that it results in a minimum cost. Pseudo code for estimation algorithm is as follows:

for i 1 to n do
for j 0 to n do

. for app
. for core
if (f_{min}[i] > f_{core}[j]) then
costmatrix[i][j] 1;
else
execute app[i] on core j.
Get execution time and cost
if execution time > timing constraint then
cost[i][j] 1 ;
else if power budget > max power constraint for chip then cost[i][j] 1 ;
else
update cost[i][j] to estimated cost value

Apply Allocation algorithm on cost matrix. Display output as one to one core allocation.

4 Adaptive Application Mapping Architecture

Fig.-1 shows the ow of data from one component to another in adaptive ap-plication allocation algorithm. Initially data was obtained from sniper using its test application FFT with di erent split radices. This data is stored in a text le and will be fed to two components namely allocator/manager and estima-tion engine. Estimation engine estimates minimum frequency requirements for each application and estimate the cost using previous cost values and actual cost which it gets from a component called Cores. It also saves the actual cost values in an array and text le. After estimating the cost, it scales it to an integer value which is then fed to allocator along with a minimum frequency requirements of each application. Allocator checks whether a particular core can be used for an application and if it cant be used then it multiplies the cost with a large number. It then uses an allocation algorithm (Munkres used in this case) which gives an optimal mapping i.e., a suitable core for each application which is fed to cores so that the mapped core can be used for the application it has been mapped with. Cores then output the actual cost values for the cores which are used for a particular value and send it to estimation engine so that these values can be used for estimation of costs in future.

4 Akash Yadav , Mushtaq Ahmed, and Daitule Omkar Vilas

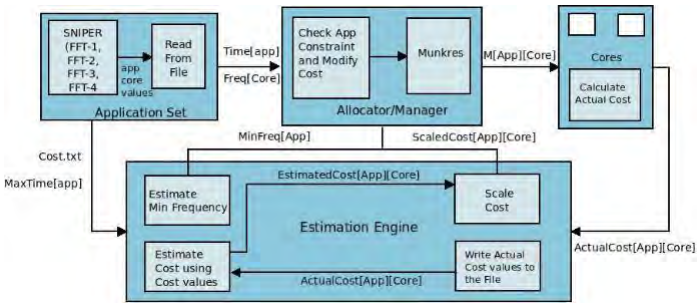


Fig. 1. Adaptive Application Mapping Architecture

IV. EXPERIMENTAL SETUP

A test application FFT of sniper simulator with di erent split radices are used as independent processes. Sniper is a parallel, high-speed multi-core simulator

[9]. These applications are compiled and then scheduled on di erent core fre-quencies like 2 GHz, 2.5 GHz, 3 Ghz and 3.66 GHz. The total execution time of an application to run on core, power leakage values and runtime dynamics are estimated with the help of McPAT framework [10] and stored in 2D cost matrix. This cost matrix is fed to allocation algorithm which gives an optimal mapping i.e., a suitable core for each application to

run. The timing analysis and the power analysis are manipulated which is shown in the Fig.2.

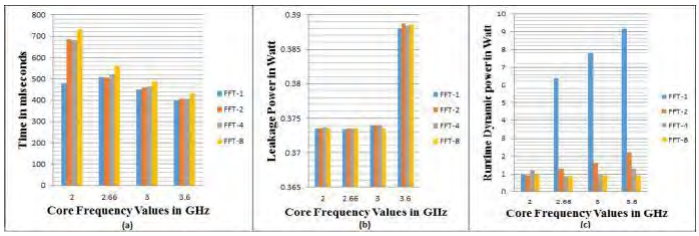


Fig. 2. (a) Total execution time for processes on di erent core frequencies (b) Leakage power values for processes on di erent core frequencies (c) Runtime Dynamic power values for processes on di erent core frequencies

Table 1. Cost matrix

Application	Core-1 (2 GHz)	Core-2 (2.66 GHz)	Core-3 (3 GHz)	Core-4 (3.66 GHz)
FFT-1	1.0035	6.36631	7.79809	9.1548
FFT-2	0.940722	1.31966	1.619870	2.1909
FFT-4	1.20623	0.916404	0.980522	1.2567
FFT-8	0.967971	0.898728	0.927073	3.9912

VII RESULTS AND ANALYSIS

This implementation is quite exible allowing the user to use di erent estimation parameters as per need. Here we have taken runtime dynamics as a parameter to estimate cost for running an application on a core. Table-2 given below contains di erent cost values when the input data, mentioned in section-5, was fed to adaptive allocation algorithm allocating cores in the manner given in the table and when same data fed to the simple allocation algorithm (if applications are randomly mapped on core available at that time).

Table 2. E ective cost by di erent method

Mapping of processes by Simple available core allocation method			Mapping of processes by Adaptive allocation method		
Application	Allocated Core	Cost	Application	Allocated Core	Cost
FFT-1	Core 1	7.294882	FFT-1	Core 1	4.506933
FFT-2	Core 2		FFT-2	Core 2	
FFT-4	Core 3		FFT-4	Core 3	
FFT-8	Core 4		FFT-8	Core 4	

Cost value taken by simple allocation is almost two times the value obtained when adaptive allocation algorithm is used. The results will be better if one of the highest costs is along the diagonal in the cost matrix because in that case simple allocation will take that cost. In some cases, simple allocation from available cores may result in lesser cost than the optimal mapping. This anomaly can be due to the fact that simple allocation overlooks the timing and frequency constraints of an application. Thus, if this algorithm is used then we can use the process exploitation to get better results and at the same time better cores can be used minimizing the overall cost of running a complete application.

V. CONCLUSIONS AND FUTURE WORK

Results of proposed work show that algorithm results in an optimum mapping of applications with given cores which helps us in achieving our goal of getting the overall minimum cost. This algorithm can be extended for dynamic cores which can operate in more than one frequency con gurations. This can also be used on a large scale once if it can be implemented on a chip level and thus can be used by wide variety of applications i.e. extending its scope from simple video coding to other applications which are a ected by process variation and dark silicon. Thus, if an application is able to select better core from a pool of available cores then it can e ectively reduce its running time and thus giving better results by consuming less power.

References

1. Shekhar Borkar, Tanay Karnik, Siva Narendra, Jim Tschanz, Ali Keshavarzi, and Vivek De. Parameter variations and impact on circuits and microarchitecture. In Proceedings of the 40th Annual Design Automation Conference, DAC '03, pages 338{342, New York, NY, USA, 2003. ACM.
2. Bharathwaj Raghunathan, Yatish Turakhia, Siddharth Garg, and Diana Mar-culescu. Cherry-picking: Exploiting process variations in dark-silicon homogeneous chip multi-processors. In Proceedings of the Conference on Design, Automation and Test in Europe, DATE '13, pages 39{44, San Jose, CA, USA, 2013. EDA Consor-tium.
3. Santiago Pagani, Heba Khdr, Waqaas Munawar, Jian-Jia Chen, Muhammad Sha que, Minming Li, and Jorg Henkel. Tsp: Thermal safe power: E cient power budgeting for many-core systems in dark silicon. In Proceedings of the 2014 Interna-tional Conference on Hardware/Software Codesign and System Synthesis, CODES '14, pages 10:1{10:10, New York, NY, USA, 2014. ACM.
4. Heba Khdr, Santiago Pagani, Muhammad Sha que, and Jrg Henkel. Chapter four - dark silicon aware resource management for many-core systems. In Ali R. Hur-son and Hamid Sarbazi-Azad, editors, Dark Silicon and Future On-chip Systems, volume 110 of Advances in Computers, pages 127 { 170. Elsevier, 2018.

5. Anil Kanduri, Amir M. Rahmani, Pasi Liljeberg, Ahmed Hemani, Axel Jantsch, and Hannu Tenhunen. A Perspective on Dark Silicon, pages 3{20. Springer Inter-national Publishing, Cham, 2017.

6. S. Maiti, N. Kapadia, and S. Pasricha. Process variation aware dynamic power management in multicore systems with extended range voltage/frequency scaling. In 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), pages 1{4, Aug 2015.

7. S. Dighe, S. R. Vangal, P. Aseron, S. Kumar, T. Jacob, K. A. Bowman, J. Howard, J. Tschanz, V. Erraguntla, N. Borkar, V. K. De, and S. Borkar. Within-die variation-aware dynamic-voltage-frequency-scaling with optimal core allocation and thread hopping for the 80-core tera ops processor. IEEE Journal of Solid-State Circuits, 46(1):184{193, Jan 2011.

8. James Munkres. Algorithms for the assignment and transportation problems. Jour-nal of the Society for Industrial and Applied Mathematics, 5(1):32{38, 1957.

9. Trevor E. Carlson, Wim Heirman, and Lieven Eeckhout. Sniper: Exploring the level of abstraction for scalable and accurate parallel multi-core simulation. In Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, SC '11, pages 52:1{52:12, New York, NY, USA, 2011. ACM.

10. Sheng Li, Jung Ho Ahn, Richard D. Strong, Jay B. Brockman, Dean M. Tullsen, and Norman P. Jouppi. Mcpat: An integrated power, area, and timing modeling framework for multicore and manycore architectures. In Proceedings of the 42Nd Annual IEEE/ACM International Symposium on Microarchitecture, MICRO 42, pages 469{480, New York, NY, USA, 2009. ACM.

Fraud Prevention on Crowdfunding Platforms

Megha Sahu[1],Ashish Gangaramani[2] ,Asha Bharambe[3]
[1] BE I.T VESIT,Mumbai University,Mumbai,India
[2]BE I.T VESIT Mumbai University,Mumbai,India
[3]Associate professor at VESIT Mumbai University,Mumbai,India

Abstract.**This paper proposes a solution on how to prevent fraud in current crowdfunding platforms using blockchain and smart contracts.Crowdfunding has some bad effects of an evil world. The main aim of this solution is to propose a solution that can reduce those effects. The key feature of Blockchain is that it maintains transparency among the nodes in the network. Blockchain-based smart contracts are contracts that can be partially or fully executed or enforced without human interaction. Since smart contracts are fundamentally conditional transfers of values, economic mechanisms designed in the game theory literature they can be implemented as smart contracts designed for crowdfunding websites where campaign managers can get approval based on their requirements. In this paper, we are proposing the methodology of how some key features of blockchain makes it efficient to implement crowdfunding.**

Keywords. Crowdfunding, Smart Contracts, Blockchain

I. INTRODUCTION

Crowdfunding is a method of raising capital through the collective effort of friends, family, customers, and individual investors. This approach taps into the collective efforts of a large pool of individuals—primarily online via social media and crowdfunding platform and leverage their networks for greater reach and exposure.

1.1 Benefits of Crowdfunding

From tapping into a wider investor pool to enjoying more flexible fundraising options, there are a number of benefits to crowdfunding over traditional methods. Here are just a few of the many possible advantages

Reach – By using a crowdfunding platform you have access to thousands of accredited investors who can see, interact with, and share your fundraising campaign.

Presentation – By creating a crowdfunding campaign, you go through the invaluable process of looking at your business from the top level—its history, traction, offerings, addressable market, value proposition, and more—and boiling it down into a polished, easily digestible package.

PR & Marketing – From launch to close, you can share and promote your campaign through social media, email newsletters, and other online marketing tactics. As you and other media outlets cover the progress of your fundraise, you can double down by steering traffic to your website and other company resources.

Validation of Concept – Presenting your concept or business to the masses affords an excellent opportunity to validate and refine your offering. As potential investors begin to express interest and ask questions, you'll quickly see if there's something missing that would make them more likely to buy in.

Efficiency – One of the best things about online crowdfunding is its ability to centralize and streamline your fundraising efforts. By building a single, comprehensive profile to which we can funnel all our prospects and potential investors, we eliminate the need to pursue each of them individually.

1.2 Problem Statement

"Nicole Carpenter"[7] have stated that current crowdsourcing platforms are at high risk by malicious campaign creators who may raise money in the name of a project and inappropriately misuse funds for their private use reason for the same,this is due to lack of transparency between investors and campaign managers.Our main aim is to create a platform which maintains transparency on how campaign manager will use the raised funds.Blockchain will help us to forecome the problem of transparency between investors and campaign manager.

Under the current structure, the ultimate amount raised for a crowdfunding campaign is almost never the actual amount raised. That's because middlemen clogg the funding pipeline, giving rise to the phenomenon of a middleman.Blockchain technology automated processes, from smart contracts to security features and more, do not call for additional fees, nor does the technology distinguish between these services. Ultimately, this provides the opportunity for the technology to be a cheaper, one-stop platform that could replace most, if not all, of the intermediaries

siphoning money from crowdfunding campaigns under the current structures.

"CJ Cornell" and "Charles Luzar" [9] have mentioned the possible frauds on crowdfunding platform "Backer Fraud" when backer submits campaign to withhold the funds or to claim the refund."Broker/Portal Fraud" where campaign operators themselves may engage in fraud, or enable fraud."Backer Creator Fraud" where creators contributed to their own campaigns either directly or through surrogates.

1.3 Objective

- Our main objective is to create a web based solution for crowdfunding using blockchain.So that it can be easily accessed by the end user.
- Identify the possible frauds, vulnerabilities of currently implemented crowdfunding solutions.
- Provide more secure solution to the user by preventing above specified frauds to occur.

1.4 Scope

Although there are many problems and frauds related to crowdfunding But our scope is limited to few major section of any crowdfunded project

- **Preempted Fraud**-In "Preempted Fraud," a suspicious crowdfunding campaign is shut down by the platform before any money changed hands. The policing is often done by members of the site (many who are just observers as opposed to contributors).
- **Attempted Fraud**-Attempted fraud could take place due to a campaign using IP that they did not own, or similar misrepresentations. While these deliberate attempts often are 'preempted' or 'stillborn', several campaigns have been launched that have used copyrighted IP or have disputable claims.
- **Perceived Fraud**-In these cases, accusations of fraud often come when the 'perks and rewards' are significantly delayed. Rewards-based crowdfunding has spawned a new category called 'pre-order' or 'pretail' crowdfunding – where backers get claim to an early version of the new product they are backing as a perk in exchange for contributing to the project.

II. REVIEW OF LITERATURE

2.1 Papers/Findings

1. **The paper titled "Exploring Blockchain for Alternative Finance" stated that[1]:**
   This paper was based on the variety of research on crowdfunding and its impact on society and economy.They created a work group to explore blockchain for alternative finance.Their main aim was to enable different target groups of the sector, namely crowdfunding platforms, investors, startups, SMEs, and policymakers to understand, deploy and support blockchain empowered use cases.The basis of their research were qualitative interviews with market and technology experts and SMEs, who were experienced in blockchain based products/services. Based on their outcomes, they complemented the practical perspectives from a research, legal and technological point of view.They did research keeping in mind - technical aspect,particality ,legal aspect.They stated that blockchain can be a solution to various security issues like confidentiality , fraud and trust.
2. **The paper titled Funding Community Projects with Smart Contracts on Blockchain[2] stated that :**

   The main aim of this paper was to demonstrate free riding problem in civic crowdfunding i.e people can still take advantage of projects without investing in it.This could be a reason for no investment in the campaign.Hence they suggested a way to implement incentive mechanism with cryptocurrency like Ether.They also stated that as blockchain is distributed the interaction between the agent and the smart contract will be guaranteed to be recorded in the blockchain.Since each agent in the blockchain is identified only by its public key, mechanisms deployed as smart contracts are vulnerable to Sybil attacks. And as Solidity programming language and the Ethereum execution environment are still in infancy, several work-arounds are needed to deal with floating points, buffer over/under flows, reentrant code etc.

3. **The paper titled "Equity Crowdfunding Based on the Blockchain? – A Delphi Study [3]" stated that :** The purpose of the study was to conclude that equity crowdfunding using blockchain has potential to close the gap faced by startups.They divided this study into two

parts first expert based delphi survey and ten market driving forces and checking how they influence the equity crowdfunding if

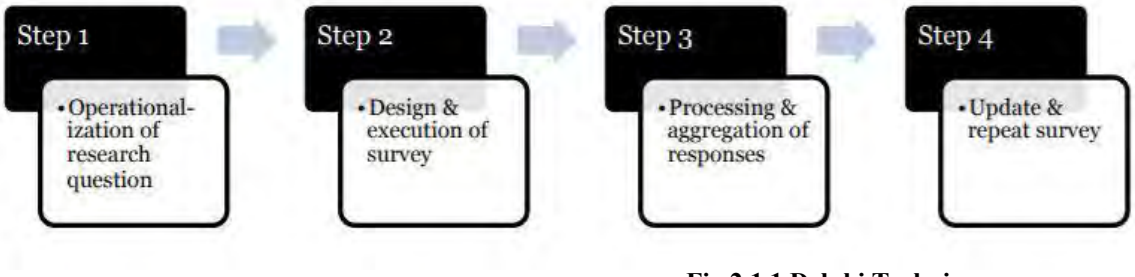


Fig 2.1.1 Delphi Technique

Fig 2.1.1 illustrates classic delphi techniques,which starts with operationalization of research question based on literature review and expert opinions that can be evaluated by further experts.Then, a survey is designed and executed with a pilot round beforehand. After conducting the first round, anonymous responses are processed and statistically aggregated for the next round. This controlled feedback allows experts to reassess their initial judgements. The whole process is repeated until a previously determined stopping criterion is reached.The result of study showed that out of 10 driving forces, 6 had positive influence on equity crowdfunding.And hence experts agreed that blockchain has positive influence on equity crowdfunding.

4. **The paper titled "Analysis and outlook of applications of blockchain technology to equity crowdfunding in China.[4]" stated that :**

This paper was specifically for crowdfunding in china.They reviewed different papers and already implemented applications to conclude if blockchain can be used in equity crowdfunding or not. They wanted to allow anyone with a good startup idea to create campaign.They stated that implementing the solution with blockchain will provide low barrier entry points,low cost and high speed transaction as compared to traditional application.Their main aim was to maintain transparency and fight money laundering.As the reward to the investor,they offer equity or bond like shares.They have also mentioned some problem that is faced in implementing equity crowdfunding "Registration of shares - the location of investors can be drastically different" , "Difficulty in confirmation of shareholder's rights "."Registration cost management- different location different cost". Blockchain for equity is still on paper,there are many technical and legal issues to be resolved.

5. **Whirl White paper "Inspiring a Whirlwind of Good[5]" stated that :**

This is a whitepaper on an application called WHIRL "Pay-it-forward" which an crowdfunding project based on blockchain.The most interesting aspect in their application is that whenever someone creates a campaign, they will get karma points depending on the amount you contribute. A person can create

campaign only when they have some karma points i.e they have contributed to some campaign.As there can be lots of campaign this application only focuses on limited campaign at a time combined with karma points.And their unique point is that they use different dozen of cryptocurrencies hence this project can be used globally.

They are tackling two problems first "Many campaign closes unfunded" and secondly backers loses interest when they don't get their rewards.To solve first they have limited campaign at a time and only people long listed on different cryptocurrencies can back a project.And they give karma points in reward.

6. **The paper titled "The Application of Blockchain Technology in Crowdfunding:**

Towards Financial Inclusion via Technology" [6] stated that :

This paper is on crowdfunding in Malaysia, which is the first country in south asia to implement it.They stated that "Crowdfunding is a practice of funding a project or venture by raising small amounts of money from a large number of people via the internet". According to paper crowdfunding can be a different financial source for startups ,entrepreneur .Blockchain could bring crowdfunding to the next level by providing not only data security but also efficiency and affordability.Building awareness is most important in crowdfunding.Creating roadshows,conferences are necessary in empowering underprivileged groups.They tried to create engagement with private sectors and also created shariah - complaint crowdfunding to encourage crowdfunding in muslim countries.

III. PROPOSED SOLUTION

Author "Conny Weber and Christin Friedrich"[1] and " Praphul Chandra, Amit Ranjan, Jaywardhan Sawale, Lakshmi Rajsekhar,Gaurang Singh, and Hrishi Wadki" [3] have concluded that the blockchain has good influence on crowdfunding and can be beneficial for society and economy of the country if implemented properly.We can conclude that providing incentives to the investor is a great aspect in crowdfunding. "Analysis and outlook of applications of blockchain technology to equity crowdfunding in China" [2] and "Inspiring a Whirlwind of Good " [5] have implemented different ways of providing incentives to the investor.But we find that giving ether as incentives will be more beneficial for investor than giving karma points."Aishath Muneeza,Nur Aishah Arshad,Asma' Tajul Arifin" [6] have stated how malaysia has implemented crowdfunding using blockchain ,how creating awareness about crowdfunding is important and that needs support of financial banks also.

An important aspect in crowdfunding is security and management of the money funded by the investors,hence we will be using Blockchain to maintain transparency between investor and campaign creator.The investors will have all the right to decide if the campaign creator can withdraw the money that have invested or not after threshold reached.

We are using ethereum as it is public blockchain network that allows anyone to get connected to the network.No node in ethereum will have special rights like in other type of network.Ethereum uses proof of Stake.While PoW depends on computing power to mine blocks, PoS depends on a node's stake (typically the amount of currency a user holds) in the system. The more stake a user controls, the more authority they have over validation.Hence this is more suitable for crowdfunding as it gives more priority to higher investors.

3.1 Methodology

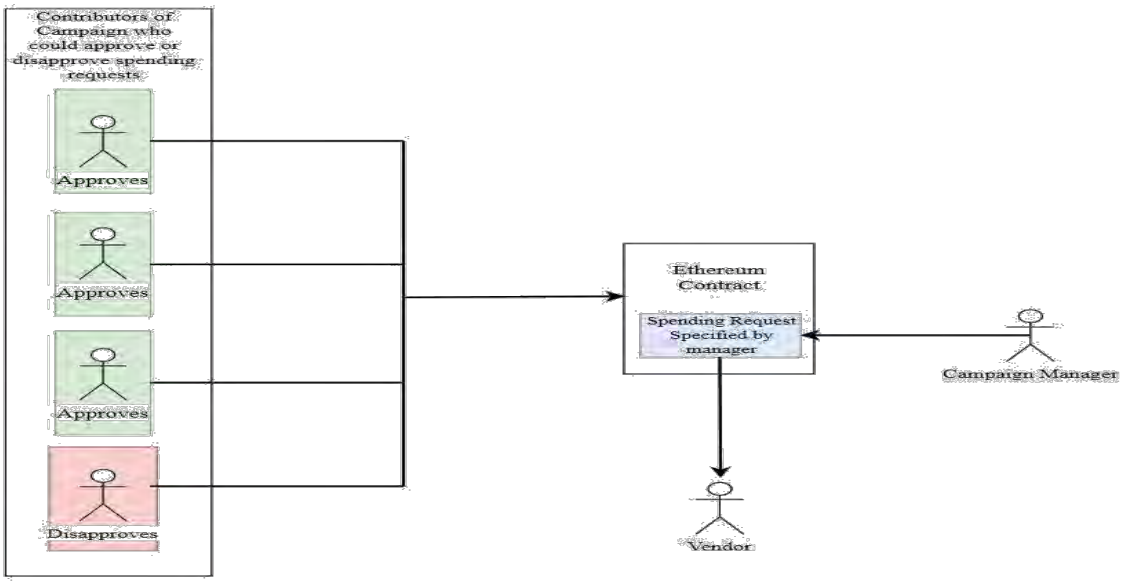


Fig 3.1.1:The Solution Architecture

Fig 3.1.1 illustrates that manager of the campaign will create a spending request on an Ethereum smart contract contributors will approve the spending request on platform.

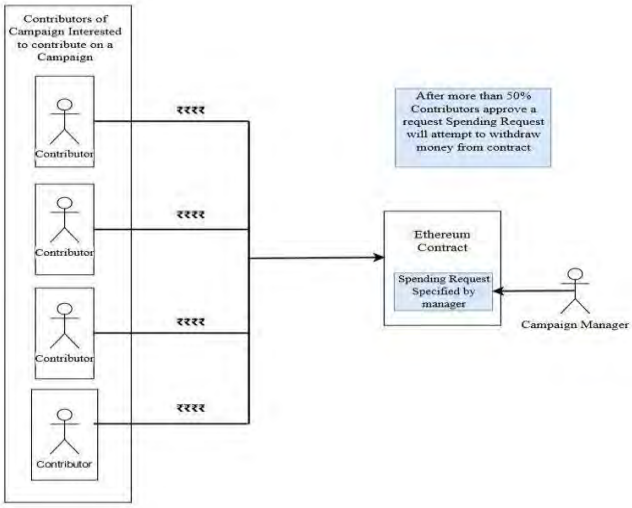


Fig 3.1.2: An Approval Module

In Fig 3.1.2 green color represents approval of backers as YES and red color as NO and depending on approval results the contributors will contribute accordingly. The money will be given to vendor and vendor will provide the requirements to campaign creators. Contributors in return will get some credits or offers as they contributed towards the campaign In this way campaign creators cannot fake their requirements and use it for their personal use This way campaign would be successful and technological world will be benefited from the innovations created by campaigns.

End User Diagram

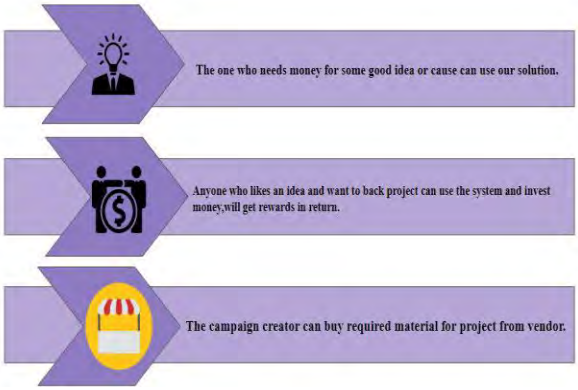


Fig: 3.1.3: Users of the system

Fig no 3.1.3 illustrates all the users of the application i.e campaign manager,investors and vendor.These have to be connected to the blockchain,to get connected they can register using application.

Use Case Diagram

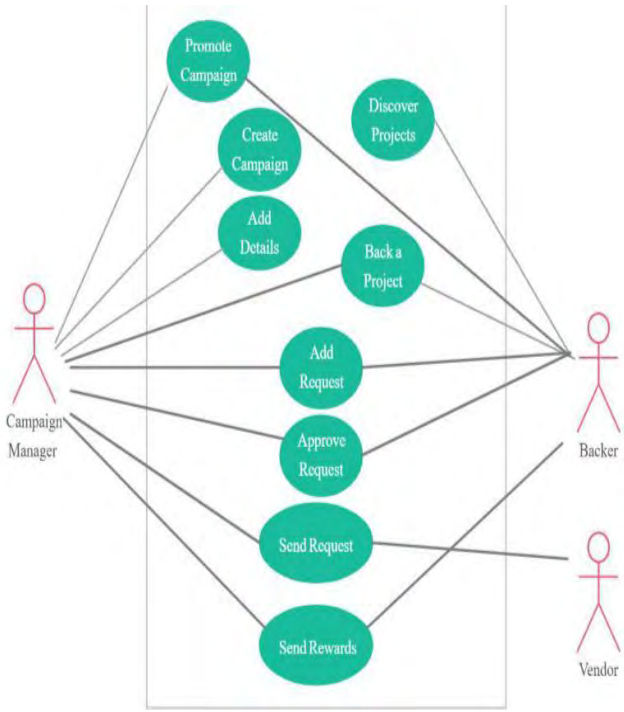


Fig 3.1.4 Use case diagram

Fig 3.1.4 shows the use case diagram of the application which illustrates the involvement of different end user in activities.Example for a campaign manager to add request there must be some investors.

IV. IMPLEMENTATION AND RESULTS

A web based application is developed for our solution and to handle block chain side of our application we have used solidity programming language, for end users to interact with our smart contract we have used React JS as our front end and to provide an interfacing between front end and block chain web.js is used as a translation layer.

4.1 Rinkeby

Rinkeby provides us with an Ethereum testnet it is used by developers to test their written smart contracts. Currency on Rinkeby testnet is valueless. Rinkeby uses Proof of Work. We have used Rinkeby test network to

make sure that all transactions are working in perfect manner when they will be deployed to an actual ethereum network, at that time there should not be a risk involved of users losing their cryptocurrency which possess an actual value.

4.2 MetaMask

MetaMask is a browser extension which enables us to run decentralized applications (dApps) without being part of the Ethereum network as a Ethereum node. Instead it connects to another ethereum node called Infura and run smart contracts on that node. MetaMask manages Ethereum wallet, which contains Ethers of a user, and allow us to send and receive through a decentralized application of interest.

4.3 Pages of application

These are some prototype design of our crowdfunding solution on blockchain, it contains a series of webpages that a normal end user will use to interact with an instance of a contract deployed on rinkeby test network.

Campaign Display page

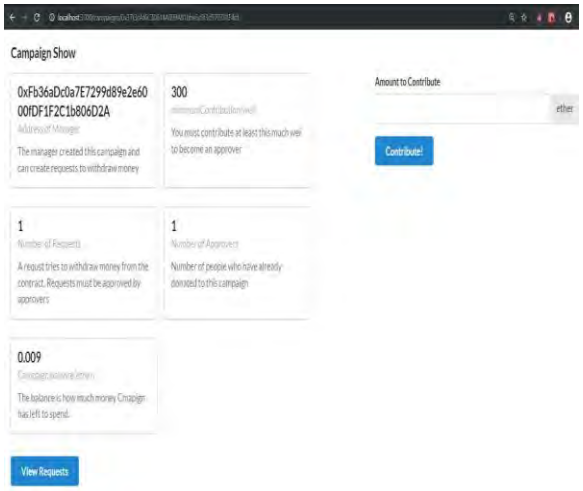


Fig: 4.3.1 View Campaign page

The new campaign has been added to the application and Fig 4.3.1 displays that campaign,on this page users can see what the campaign is about and various details about campaign like minimum contributions required, number of requests and number of approvers who have approved for this campaign. Users could also contribute to that campaign on this page. To view all requests that campaign manager has requested view requests button can be clicked.

Requests page

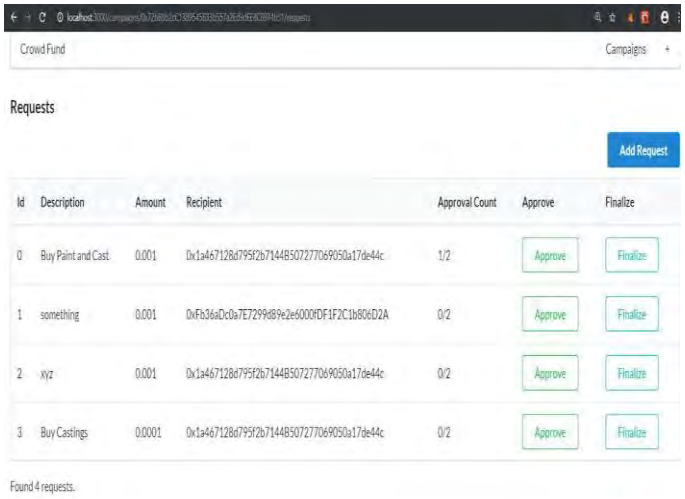


Fig: 4.3.2 View Requests page

Fig 4.3.2 shows how users could view all the requests created by campaign manager, also they can approve the particular request but only if they have contributed to that campaign that request can only be finalized by campaign manager only and only if percentage of requests to campaign contributors is greater than 50%.

4.4 Etherscan

Etherscan is a blockchain explorer for Ethereum and enables us to search the Ethereum blockchain for transactions, addresses, tokens, prices and other activities that take place on Ethereum or on Rinkeby test network. We are using Etherscan to confirm details of what is published on Rinkeby test network.

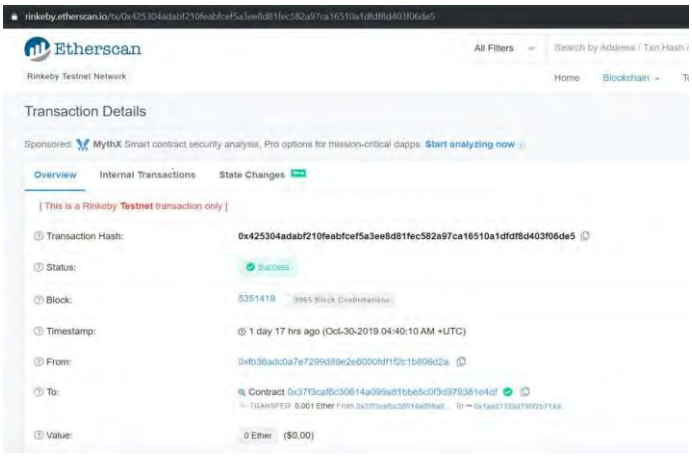


Fig 4.4.1 Details of a Transaction on Etherscan

Fig 4.4.1 shows transaction details on etherscan we can search details about any transaction by simply entering its hash no matter how old a transaction is we could verify that transaction on etherscan till we have its hash.Ether scan shows the following details:

- Status of a transaction weather it was successful or not.
- Block number of that transaction on Blockchain.
- Timestamp when it was mined, From whom this transaction was sent its Public address.
- To whom this transaction was intended to.
- Gas used for processing of this transaction.
- Nonce Position of the block in Blockchain.

All the above Transaction details ensures transparency as its publicly available and these details further cannot be modified.

V. CONCLUSION

The execution environment for Blockchain based Smart Contracts is the peer-to-peer network of thousands of geographically distributed nodes which maintain a copy of the blockchain (distributed ledger).In a blockchain each node/agent is identified by its public key, for testing environments Rinkeby network could be used to create demo users for a network, Ethereum and solidity programming has to be used for creation of smart contracts which will interact with our blockchain network.Later the application will be hosted on AWS EC2 and smart contract will be hosted on the main ethereum network, any person on the blockchain can access the application.The application will be beneficial for society, people will get single platform to showcase their innovative ideas,people who needs money in some crisis can also use this app.Donator can trust this solution as campaign creator won't be able to use money other than the cause.

References

[1]Conny Weber and Christin Friedrich **"**Exploring Blockchain for Alternative Finance**"** A publication of the ECN working group scoping paper.

[2] Huasheng Zhu and Zach Zhizhong Zhou **"**Analysis and outlook of applications of blockchain technology to equity crowdfunding in China**"** Zhu and Zhou Financial Innovation 2016.

[3] Praphul Chandra, Amit Ranjan, Jaywardhan Sawale, Lakshmi Rajsekhar,Gaurang Singh, and Hrishi Wadki **"**Funding Community Projects with Smart Contracts on Blockchain**".**

[4] Felix Heieck , Tatiana Ermakova Benjamin Fabian and Stefan Lessmann **"**Crowdfunding Based on the Blockchain? – A Delphi Study**"**Humboldt University Berlin, Chair of Information Systems.
[5] Whirl **"**Inspiring a Whirlwind of Good**"** White paper.
[6] Aishath Muneeza,Nur Aishah Arshad,Asma' Tajul Arifin **"**The Application of Blockchain Technology in Crowdfunding: Towards Financial Inclusion via Technology paper**"** International Journal of Management and Applied Research, 2018, Vol. 5, No. 2.

[7] https://www.digitaltrends.com/cool-tech/biggest-kickstarter-and-indiegogo-scams/ last accessed Date:28/11/2019

[8]https://www.forbes.com/sites/devinthorpe/2017/11/29/worries-about-fraud-top-list-of-crowdfunding-concerns/#1 94d842c48f0last accessed on Date: 28/11/2019
[9]https://www.crowdfundinsider.com/2014/03/34255-crowdfunding-fraud-big-threat/

last accessed Date: 28/11/2019

Review of Wireless Sensor Network Security Schemes

Vaibhav Dabhade[1] and Dr. A. S. Alvi[2]

Computer Engineering, MET´s BKC, IOE, Nashik, Maharashtra, India I.
vaibhavdabhade@rocketmail.com, home page: https://metbhujbalknowledgecity.ac.in/
Information Technology, Ram Meghe Institute of Technology & Research, Badnera, Maharashtra, India

Abstract. **Wireless Sensor Network is used in various application such as hospital, environment monitoring, experiments. It is also used in bat-tlefield for target tracking. If it is deployed in hostile environment in military application then nodes can be captured and data can be up-dated. WSN is equipped with low battery, storage and computational power. Hence, how to ensure the security of wireless sensor network un-der limited resources has become a key problem. This paper starts with application of WSN, summarizes security threats and analyses various key management and security schemes.**

Keywords: wireless sensor network security, issues, key management, pairwise key generation

I. INTRODUCTION

A wireless sensor network (WSN) is a group of spatially distributed autonomous devices i.e. sensors. It can be used to monitor physical or environmental con-ditions such as sound, temperature, vibration, pressure, motion or pollutants

[14]. Military application such as battlefield surveillance [12], is the main moti-vation for the development of wireless sensor networks. However, wireless sensor networks are now used in many application areas, including industrial process monitoring and control, machine health monitoring, environment and habitat monitoring, healthcare applications, traffic control, and home automation [3]. Wireless sensor networks have emerged as a dominant technology in the current decade. WSNs are becoming popular now a days because of its applications. Al-ready many companies have started developing commercial applications. At the same time they have posed numerous unique challenges to researchers. Wireless sensor networks are generally composed of hundreds or even thousands of tiny sensor nodes, which are constrained in many aspects like memory capacity, pro-cessing power and most importantly energy or battery power. Most of the time sensor nodes do not have access to renewable energy resources. The overall cost of a WSN is also required to be relatively lower. Moreover when we look at the applications of WSNs, there are many applications areas, e.g., battlefield aware-ness, traffic monitoring system etc in which security of information remains as an important issue. Providing security to a WSN is a non-trivial problem [26]. Security mechanisms which are applicable to wired or other ad-hoc networks are not suitable for WSNs [12]. There are many reasons behind it and we discuss those in the subsequent sections. There are many challenges in sensor networks. In this paper we focus on different security issues. Wireless sensor network ap-plications include wildlife monitoring, manufacturing machinery performance monitoring, building safety, earthquake monitoring, and military applications. A major benefit of these systems is that they perform in-network processing to reduce large streams of raw data into useful aggregated information. Protecting it all is critical. Because sensor networks pose unique challenges, traditional se-curity techniques used in traditional networks cannot be applied directly [18]. Several key challenges need to be addressed in wireless sensor network. First, to make sensor networks economically viable as sensor devices are equipped with limited energy, computation and communication capabilities. Second, sensors are often deployed in accessible areas where physical attack possibility increases. And third, sensor networks interact closely with their physical environments and with people, posing new security problems [15]. Hence, existing security schemes are inadequate, and new schemes/ideas are required for security. Here, we out-line security issues in WSNs, followed by schemes for sensor network security and suggest future work for research. We cover several important security challenges, including key establishment, secrecy and authentication.

1.1 WSN ARCHITECTURE

Layerd Architechture: A layered architecture shown in 1consist of a single powerful base station and sensor nodes layers around it correspond to the nodes that have the same hop-count to the BS. Layered architectures are used in-

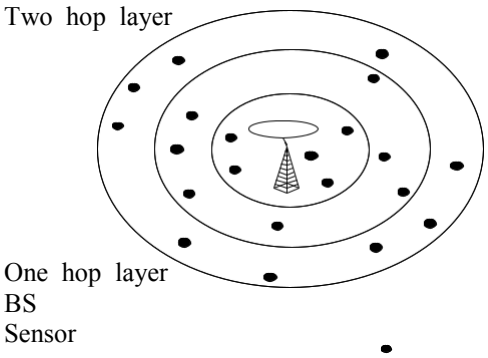


Fig. 1. Layered Architecture of WSN.

building wireless backbones. The advantage is that each node is involved only in short distance.

Clustered Architechture: Clustered architecture organizes the sensor nodes into clusters. As depicted in figure 2, each cluster is governed by a cluster-head. Nodes send data to cluster head and then only heads send messages to a Base Station. This is Suitable for data fusion and it is Self-organizing.

While analyzing security schemes we need to understand WSN limitations. Lim-

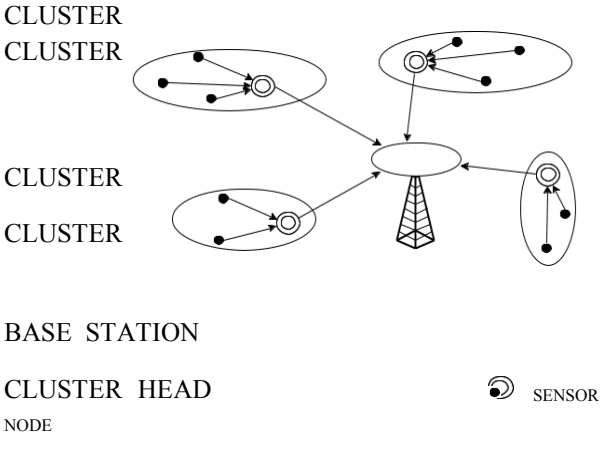


Fig. 2. Clustered Architecture of WSN.

itations of WSNs are as follows. Computational Power[3]: Symmetric and asym-metric cryptographic schemes are available for WSNs security but each schemes have their own advantages and limitation. Resilience of asymmetric schemes is good but it requires high computation and sensor nodes are equipped with very low computational power. At the same time it requires more power that symmet-ric schemes. High computations power drains battery of sensor node. Hence these issues need to be consider while proposing and implementing security mechanism for WSNs.

II. ISSUES/ATTACKES

In Wireless Sensor Networks hackers/intruders can gain access to data collected by nodes. They can get access remotely to nodes. Once they get access to node, attackers read/modify data depending on type of attack. In passive type of at-tack, attacker collects data and analysis characteristic of data flow. In active type of attacks, attackers modify data for his own benefit. Comparatively pas-sive type of attacks are difficult to detect but can be protected by encrypting the contents. Encryption protects data but because of limited resources, we need to be very careful while selecting and implementing encryption algorithms.

The term Security is mainly comprises of authentication, confidentiality and in-tegrity. Authentication allows a sensor node to ensure the identity of the peer node with which it is communicating. It rejects messages coming from attackers or any other send if they fail to prove identity. Confidentiality is making in-formation inaccessible to unauthorized user. Only authorize user can get access to data. In symmetric encryption techniques same key is used to encrypt and decrypt. Cipher text is transmitted to receiver and on receiving side cipher text will be decrypted by secret key. Integrity is used to check that the received data is not altered. Message contents can be modified by attackers. If node compro-mised then data integrity is violates [6]. In addition to this we should know one more term i.e. availability. It allows authorize user to access data from anywhere and anytime as per request. Below section focuses on key management, node capturing, and addition on new node in network.

III. ANALYSIS OF SECURITY SCHEMES

Key management schemes are the core of the secure communications. Config-uring a secure sensor network is a non-trivial task due to resource constrained sensors. As we have seen, existing key management schemes for the WSN are not able to achieve security and efficiency requirement. To address weakness of existing schemes and to achieve security and efficiency requirement we propose key establishment scheme based on pairwise key approach. Pairwise key concept assures highest resilience for WSNs. This chapter deals with key management and establishment schemes, in which we explain schemes in detail. Considering security, key management is very important and complex in wireless sensor net-work. Wireless Sensor Networks dynamic structure increase the difficulty of key management. Various key distribution schemes have been proposed to establish pairwise key between nodes in the literature. Each scheme has its weakness

and cannot simultaneously achieve both security and efficiency requirements of wire-less sensor network. This section deals with the work carried out on various key management schemes for WSNs.

3.1 Key-Management Scheme by Eschenauer and Gligor

In [21], key distribution consists of three phases, 1. Key pre-distribution, 2. Dis-covery of Shared key, and 3. establishment of Path key. The key pre-distribution phase [21] of this scheme consists of five off-line steps, after key pre-distribution phase every node has to discover shared key with its neighbors. During net-work initialization in the operational environment the shared-key discovery phase takes place, where every node discovers its neighbors in wireless communication range with which it shares keys. In path-key establishment phase path-key is assigned to selected pairs of sensor nodes in communication range. Since all the keys are randomly selected from the same key pool, it is quite possible that two neighboring nodes have some overlapped keys. If common key in two sensors identified, sensors can use it as their pairwise key directly. Otherwise, a path key establishment procedure is imitated which generates a path-key between the two communicating nodes under some other intermediate node's participation. This scheme addresses security issues but power consumption, computational speed have to consider for effective implementation.

3.2 Random Key Pre-distribution Schemes

Chan [20] proposed improved scheme based on [21]. This scheme is also known as q-composite scheme. Any two neighboring sensor nodes find a single common key from their key rings to establish a secure link in the key setup phase. [20] Proposes a modification to the basic scheme where q common keys ($q \geq 2$) are needed, instead of just one. By increasing the amount of key overlap required for key-setup, the resilience of the network against node capture increases. If the amount overlap of required key increases, it becomes exponentially harder for an attacker with a given key set to break a link. However, to preserve the given probability p (probability that two neighbor nodes can set up a secure link during the key-setup phase) of two nodes sharing sufficient keys to establish a secure link, it is necessary to reduce the size of the key pool | S |. When size of pool gets reduced then attacker will get access to gain a larger sample of S by breaking fewer nodes. These two factors results in an optimal amount of key overlap. High level of resilience against node capture can be achieved using the q-composite keys scheme when number of captured nodes are small. If large numbers of sensors nodes have been compromised, the q-composite keys schemes tend to reveal larger fractions of the network to the adversary. If q increased it becomes difficult for an adversary to obtain small amounts of initial information from the network if he able to capture small number of initial node. In other words, the adversary needs to capture more sensors in [20] than in [21] to crack the same fraction of communications between noncaptured nodes.

3.3 Enhanced Pairwise Key Establishment Scheme for WSN

In [21] and [20] the communication pairwise keys between sensors use the preloaded keys directly. Once some nodes are captured, the adversary may crack other nodes or even the entire network. In [21], [20] and [13] high resilience and full network connectivity cannot be achieved. Due to the heavy key storage overhead of each sensor, these schemes cannot be used for a large-scale WSN. To over-come these problems [18] proposed Enhanced pairwise key establishment scheme (called as pairwise key scheme), to achieve security and efficiency requirement of WSNs. In this scheme two type of keys, set up key and pairwise key are used. We have focused on unique pairwise key generation for WSN. If adversary captured a node and exposed pairwise keys then he cannot use captured keys to find keys of other nodes because pairwise key calculated by pair of node are not repeated and can be used by respective pair only. This technique provides highest resilience, full network connectivity and low communication overhead.

3.4 Scheme based on RSA & ECC

Faleh Alfaleh and other researchers have studied and compared two popular cryptographic schemes and found it useful as compared to other cryptosystems [6]. Among all asymmetric algorithms RSA is secure and reliable algorithm. In addition to this elliptic curve cryptography (ECC) algorithm can also be used. No doubt that RSA is more secure that ECC but considering resource constrained nature of WSNs and when we apply various evaluation criteria it is clear that ECC is suitable for WSNs. The computation cost, less memory, low bandwidth and running time of cryptographic operations must be managed to increase the efficiency of WSN.

3.5 Session Key Establishment Scheme

In WSNs gateway is used to access information collected by sensor nodes. For secure communication session key needs to be exchanged between communi-cating nodes. [7] Proposed key agreement scheme. To login in the system and establishes a secure connection with the sensor node via gateway node the user makes use of his biometric feature. It restrict the password guessing attack and forgery attack. Work is verified using BAN-logic to ensure the restriction of mu-tual authentication, and replay attack. In BAN-logic participating parties while communication undergo the verification of the transmitted message. An authen-ticated user and sensor agree upon shared session key. Agreement of key is done when scheme is executed and user is verified. With the adoptability of AVISPA, which is a suitable formal security method to prove this scheme resists all the security attributes. Results shows that, this scheme defends most of attacks and communication cost is average which needs to be improved.

3.6 Mixed Public and Secret-key Cryptography

Most of the researchers recommend symmetric cryptographic methods for mak-ing Wireless Sensor Networks secure. In WSNs, key management plays an impor-tant role in security. It has to be analyzed, verified, how the given security scheme affects security when we try to keep low battery consumption, low bandwidth and high computational speed. As per literature survey it has been observed that symmetric schemes are preferred but they are not optimal at every time. Asym-metric schemes are comparatively secure but reduce performance of nodes which affects whole network. As proposed in [2] symmetric and asymmetric methods can be used by for WSNs security. In this scheme MAC is used to generate digital signature. Evaluation shows that proposed scheme gives maximum connectivity, but only if the neighbor discovery phase is long enough. It also provides good resilience after initialization phase.

IV. ANALYSIS OF SCHEMES

This section gives the overview of various schemes proposed/implementated for Wireless Sensor Network Security. Schemes considered for security are mainly based on key management concept. Table 1 shows details of various schemes and impact.

Table 1. Comparison of WSNs schemes

Protocol/Scheme	Resilience	Remark
KM Scheme by Eschenauer and Gligor[21]	High	Scheme enables key revocation but power consumption need to be reviewed.
q-composite Scheme[20]	High	The q-composite keys scheme offers greater resilience against node capture when the number of nodes captured is small.
Enhanced Pairwise Key Establishment Scheme[18]	High	Communication overhead can be reduced to improve power consumption.
SKG Protocol[1]	Good	If base station compromised then network can be compromised
RSA & ECC[6]	High	ECC better that RSA. The computation cost, less memory, low bandwidth must be managed to in-

		crease the efficiency of WSN.
Session Key Establishment Scheme[7]	Good	Results shows that, this scheme defends most of at-tacks and communication cost is average which needs to be improved.
An Efficient Key Distribu-tion Mechanism[5]	High	Memory usage and communication overhead is high which have to be address in resource constrained WSNs.
Mixed Public and Secret-key Cryptography[2]	High	In this scheme MAC is used to generate digital sig-nature.
Energy-Efficient Key Man-agement Protocols[4]	High	Scheme is efficient only for low new node addition and authentication rate.

V. CONCLUSION

In this paper we have studied WSNs architecture, its advantages, application, challenges and few methods to implement WSNs effectively. While developing system for any field/area to address problem, resource constrained nature of WSNs need to be considered. Most challenging task to design application/system with low power, communication limitation and computational capability. Lim-ited battery and processing capability restricts researchers in making highly se-cure network. If we want to achieve higher level of security then such application may consume more power. Providing security in wireless sensor network is a non trivial task. Various schemes have been proposed in literature to address problem like security, effective utilization of power, speedy computation/processing with easy implementation. To make WSNs secure key management scheme must be simple and flexible. Most of the key management schemes are based on the ran-dom key pre-distribution concept. After analyzing schemes we can conclude that, each scheme has its weakness and can not simultaneously achieve both security and efficiency requirements of wireless sensor networks. Considering current ad-vances WSNs need schemes which will deliver desirable outcome.

References

1. Manar Hamza Bashaa, Saif Mahmood Al-Alak, Ali Kadhum Idrees, "Secret Key Generation in Wireless Sensor Network Using Public Key Encryption", *In Proceed-ings of the International Conference on Information and Communication Technol-ogy*, April 15 - 16, 2019.
2. Mattia Griotti, Filippo Gandino, Maurizio Rebaudengo, "Mixed Public and Secret-key Cryptography for Wireless Sensor Networks", *Published in Tenth International Conference on Mobile Computing and Ubiquitous Network (ICMU)*, 2017.
3. Haomeng Xie, Zheng Yan, "Data Collection for Security Measurement in Wireless Sensor Networks: A Survey", *In IEEE INTERNET OF THINGS JOURNAL*, April 2019 .
4. Laurin Doerr, Michael Heigl, Dalibor Fiala, Martin Schramm, "Comparison of Energy-Efficient Key Management Protocols for Wireless Sensor Networks" *In Pro-ceedings of the 2019 International Electronics Communication Conference*, July 2019.
5. Iman Gholizadeh, Esmaeil Amiri, Reza Javidan, "An Efficient Key Distribution Mechanism for Large Scale Hierarchical Wireless Sensor Networks", *In IEEE SEN-SORS JOURNAL*, Jan 2018.
6. FALEH ALFALEH, HAITHAM ALFEHAID, MOHAMMED ALANZY, SALIM ELKHEDIRI, "Wireless Sensor Networks Security: Case study" *Published in 2nd International Conference on Computer Applications & Information Security (IC-CAIS)*, May 2019.
7. Majid Alotaibi, "An enhanced symmetric cryptosystem and biometric-based anony-mous user authentication and session key establishment scheme for WSN" *Published in IEEE Access*, Nov 2018.
8. Sonali Pandey,Rakesh Kumar, "Study of Routing Protocol Using Key Management in Wireless Sensor Network", *In Proceedings of the Second International Conference on Computing Methodologies and Communication (ICCMC 2018)*, Feb 2018.
9. Zhi Zhang, Xianbin Wen, Haixia Xu, and Liming Yuan, "Sensing Nodes Selective Fusion Scheme of Spectrum Sensing in Spectrum-Heterogeneous Cognitive Wireless Sensor Networks", *In IEEE SENSORS JOURNAL*, Jan 2018.
10. D.Sridhar raja, T.Vijayan, B.Kalaiselvi, "Advances and recent trends in wireless sensor network" *In International Journal of Pure and Applied Mathematics*, Volume 119 No. 7 2018.
11. Dr. Deepti Gupta, "Wireless Sensor Networks 'Future trends and Latest Research Challenges' ", *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, Mar - Apr.2015.
12. Xiangqian Chen, Kia Makki, Kang Yen, and Niki Pissinou, "Sensor Network Se-curity: A Survey", *Communications Surveys & Tutorials, IEEE* , June, 2009.

13. Sinan Emre Tasci, Ersoy Bayramoglu and Albert Levi, "Simple and Flexible Ran-dom Key Predistribution Schemes for Wireless Sensor Networks Using Deployment Knowledge", *International Conference on information security and assurance* , 2008.
14. Xun Yi Faulkner, M. Okamoto, E., "Secure Wireless Sensor Networks", *Availabil-ity, Reliability and Security, 2008. ARES 08. Third International Conference*, 4-7 March 2008 .
15. Tanveer Zia, Albert Zomaya, "Security Issues in Wireless Sensor Networks", *ic-snc, pp.40, International Conference on Systems and Networks Communication (IC-SNC'06)*, 2006.
16. Hasan Tahir, Syed Asim Ali Shah, "Wireless Sensor Networks A Security Perspec-tive", *IEEE* , 2008.
17. Michael Healy, Thomas Newe, Elfed Lewis, "Security for Wireless Sensor Net-works: A Review", *IEEE Sensors Applications Symposium New Orleans, LA, USA*, February 17-19, 2009.
18. Yi. Cheng and Dharma P. Agrawal, "Improved Pairwise Key Establishment for Wireless Sensor Networks", *IEEE*, 2007.
19. Y. Cheng, D. P. Agrawal, "Efficient Pairwise Key Establishment and Management in Static Wireless Sensor Networks", *in: Proc. of the 2nd IEEE International Con-ference on Mobile Ad-Hoc and Sensor Systems, Washington, DC*, Nov. 7-10, 2005.
20. H. Chan, A. Perrig, and D. Song, "Random key pre-distribution schemes for sensor networks" *In IEEE Symposium on Security and Privacy, Berkeley, California*, May 11-14 2003.
21. L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks", *In Proceedings of the 9th ACM conference on Computer and communi-cations security* , November 2002.
22. Wendi Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, "Energy-Efficient Communication Protocols for Wireless Microsensor Networks",*Proc. Hawaaian Int'l Conf. on Systems Science*, January 2000.
23. Chungen Xu,Yanhong Ge, "The Public Key Encryption to Improve the Security on Wireless Sensor Networks", *Second International Conference on Information and Computing Science* , February 2009.
24. David J. Malan, Matt Welsh, Michael D. Smith, "A Public-Key Infrastructure for Key Distribution in TinyOS",*Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Con-ference on* , Oct. 2004.
25. Adrian Perrig, John Stankovic and David Wagner, "Security in Wireless Sensor Networks", *Communications of the ACM, vol.47, No.6*,June 2004.
26. Cionca, V.; Newe, T.; Dadarlat, V., "Setting up secure wireless sensor networks", *Intelligent Computer Communication and Processing, 2009. ICCP 2009. IEEE 5th International Conference*, 27-29 Aug. 2009.
27. C. S. Raghavendra, "Wireless Sensor Networks" *Kluwer Academic Publishers*, 2004.
28. Mauri Kuorilehto, Mikko Kohvakka, Jukka Suhonen, "Ultra-low energy wireless sensor networks in practice: theory, realization and deployment by Mauri Kuorile-hto". *John Wiley & Sons Inc*, Feb 2008.

Security Issues On Cloud Computing

Akram.H.Shaikh Research Scholar , B.B.Meshram Professor
Department of Computer Engineering and Information Technology
Veermata Jijabai Technological Institute, Matunga, Mumbai
ahshaikh_p18@ce.vjti.ac.in , bbmeshram@ce.vjti.ac.in

Cloud computing has now become a necessary enterprise architectural component irrespective of the company size. Cloud solutions like AWS by Amazon, Google cloud by Google, Azure by Microsoft, etc. are providing plans at very competitive rates, thereby making it easier for companies to deploy virtual machines of any size, any capacity and any kind of processing & security requirements. But generally, the architectural security layers are very generic and do not cater to the company specific security requirements. Cloud Computing Services like Software as a Service (SAAS), Platform as a Service (PAAS) and Infrastructure as a Service (IAAS) are delivery model which provides software, application platform and infrastructure recourses to consumer. While providing these services cloud computing is facing security and privacy issues there is need to find security issues

This Paper describe broad literature survey on cloud security issues, layer wise classification of each issues like threats, vulnerabilities and solutions present in literature

Keyword: Cloud Computing, SAAS, PAAS, IAAS, Vulnerabilities attack

I. INTRODUCTION

During the last decade, the internet has known a great invasion that has totally changed the world. This rapid evolution has made access to the Internet an easy task with the advent of different devices like cellular phones, laptops, tablets and others that connected to the virtual world anywhere and at any time. In other hand, multitenant architecture, virtualization, remote access, the ubiquity of broadband networks and the establishment of universal standards of interoperability between software's. All of these elements had been a good step for the emergence of cloud computing in the world of distributed systems as a successor of cloud computing.

However, with all of these benefits, there are still a number of barriers, that effect the widespread adoption of cloud computing, and the most important one is security. The cloud computing system is based on the trust, what makes security and confidentiality the major issues. These and other security issues whether on the technical or on the legal plan decrease the confidence of cloud computing and its adoption. In this paper covers an overview about security issues in cloud computing.

II. CLOUD COMPUTING MODEL

According to the National Institute of Standards and Technology (NIST), cloud computing is defined as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and it is composed of three service models, deployed in four deployment models, and offers five characteristics [1] as shown in Fig

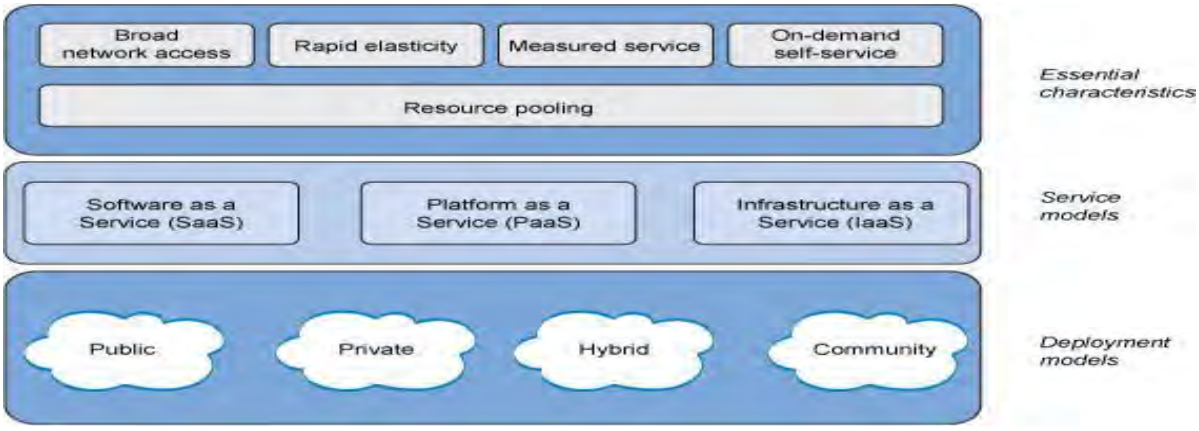


Fig: 2.1-Cloud computing services

2.1 Service Models

1] Software as a service (SaaS) The capability provided to the client is to utilize the supplier's applications running on a cloud foundation and open from different customer devices through a thin customer interface, for example, internet browser. in other words, in this model, a total application is offered to the client as a service on request.. A good example of this would be using a Web browser to view email as provided by Microsoft, Yahoo, or Google [1].

2] Platform as a service (PaaS) here the consumer can deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations [1].

3].Infrastructure as a service (IaaS) the consumer can provision processing, storage, networks, and other fundamental computing resources and he is able to deploy and run arbitrary software, which can operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure, but has control over operating systems; storage, deployed applications, and possibly limited control of select networking components [1].

2.2. Deployment Models

1. Public: the cloud infrastructure is available to the public or a large industry group and, it is owned by an organization selling cloud services.

2. Private: the cloud infrastructure operated solely for an organization. It may be managed by the organization

3. Community: the cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns. It is managed by organizations or a third party and may exist on premise or off premise.

4. Hybrid: the cloud infrastructure is a composition of two or more

2.3 Essential characteristics

1. On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically.

2. Broad network access: all services in cloud computing are available through the network and accessed by different devices such as desktop computer, mobile phones, smart phones and tablet devices...

3. Resource pooling: the provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.

4. Rapid elasticity: resources can be elastically provisioned and released. The user is free to purchase additional resources and opportunities in any quantity and at any time.

5. Measured service: Cloud systems automatically control and optimize resource use by leveraging a metering capability (pay-per-use basis) at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts).

III. CLOUD SECURITY

3.1 Security Issues on SAAS , PAAS and IAAS

This Section discusses about Security issues on SAAS, PAAS, and IAAS

3.1.1 Software-as-a-Service (SaaS) Security Issues

SaaS provides application services on demand such as email, conferencing software, and business applications such as ERP, CRM, and SCM [3]. SaaS users have less control over security among the three fundamental delivery models in the cloud. The adoption of SaaS applications may raise some security concerns.

1. **Application security**

These applications are typically delivered via the Internet through a Web browser. However, flaws in web applications may create vulnerabilities for the SaaS applications. Attackers have been using the web to compromise user's computers and perform malicious activities such as steal sensitive data. [3]. The Open Web Application Security Project (OWASP) has identified the ten most critical web applications security threats [4]. There are more

security issues, but it is a good start for securing web applications.

2. Multi-tenancy

SaaS applications can be grouped into maturity models that are determined by the following characteristics: scalability, configurability via metadata, and multi-tenancy [3]. In multi-tenancy, a single instance serves all customers. This approach enables more efficient use of the resources but scalability is limited. Since data from multiple tenants is likely to be stored in the same database, the risk of data leakage between these tenants is high. Security policies are needed to ensure that customer's data are kept separate from other customers [2].

IV. DATA SECURITY

Information security is a typical worry for any innovation, yet it turns into a significant test when SaaS clients need to depend on their suppliers for legitimate security [2]. In SaaS, authoritative information is regularly prepared in plaintext and put away in the cloud. The SaaS supplier is the one liable for the security of the information while is being handled and put away [3]

V. ACCESSIBILITY

Getting to applications over the web by means of internet browser makes access from any system device, including open PC and cell phones. In any case, it likewise opens the support of extra security dangers. The Cloud Security Alliance [5] has discharged a report that portrays the present condition of versatile registering and the top dangers here, for example, data taking versatile malware, unreliable systems (WiFi), and vulnerabilities found in the network OS and authority applications, uncertain commercial centers, and nearness based hacking

3.1.2 Platform-as-a-Service (PaaS) Security Issues

PaaS facilitates deployment of cloud-based applications without the cost of buying and maintaining the underlying hardware and software layers [2]. Same as SaaS, PaaS also brings data security issues and other challenges that are described as follows:

1. Third-party relationships

Moreover, PaaS does not only offer conventional programming languages, but also does it offer third-party web services mechanism such as mashups. Mashups merge more than one source element into a single integrated unit. Thus, PaaS models also succeed to security issues related to mashups such as data and network security [2]. Also, PaaS users have to depend on both the security of web-hosted development tools and third-party services.

2. Underlying infrastructure security

In PaaS, developers do not usually have access to the underlying layers, so providers are responsible for securing the underlying infrastructure as well as the applications services [2]. Even when developers are in control of the security of their applications, they do not have the assurance that the development environment tools provided by a PaaS provider are secure.

3.1.3 Infrastructure-as-a-Service (IaaS) Security Issues

IaaS provides a group of resources such as servers, storage, networks, and other computing resources in the form of virtualized systems, which are accessed through the Internet [1]. Users are entitled to run any software with full control and management on the resources allocated to them. IaaS providers must undertake a substantial effort to secure their systems in order to minimize these threats that result from creation, communication, monitoring, modification, and mobility. Here is some of the security issues associated to IaaS.

1. Virtualization

Virtualization enables clients to make duplicate, share, relocate, and move back virtual machines, which may enable them to run an assortment of utilizations [1]. In any case, it likewise presents new open doors for assailants in view of the additional layer that must be verified [2]. Virtual machine security becomes as significant as physical machine security, and any defect in it is possible that one may influence the other. Virtualized situations are defenseless against a wide range of assaults for ordinary frameworks; be that as it may, security is a more noteworthy test as virtualization includes more purposes of section and more interconnection multifaceted nature. In contrast to physical servers, VMs have two limits: physical and virtual [7].

2. Virtual Machine Monitor

The Virtual Machine Monitor (VMM) or hypervisor is liable for virtual machines detachment; hence, if the

VMM is undermined, its virtual machines may conceivably be undermined too. The VMM is a low-level programming that controls and screens its virtual machines, so as any conventional programming it involves security defects [1].

3. Virtual Networks

System segments are shared by various inhabitants because of asset pooling. As referenced previously, sharing assets enables assailants to dispatch cross-occupant assaults. The most secure path is to snare each VM with its host by utilizing devoted physical channels. Nonetheless, most hypervisors utilize virtual systems to interface VMs to convey all the more legitimately and effectively. For example, most virtualization stages, for example, Xen give two different ways to arrange virtual systems: spanned and steered, yet these procedures increment the likelihood to play out certain assaults, for example, sniffing and parodying virtual system [2].

VI. VULNERABILITIES THREATS AND SOLUTIONS FOR CLOUD SECURITY

Table 1 Shows Vulnerabilities Threats and Solutions for the Respective layers of the cloud.

Table 1: Layer wise Threats Vulnerabilities and Solutions

S.N	Delivery Model	Threats	Vulnerabilities	Solution or Defense Mechanism
1	SAAS PASS IAAS	Account or service Hijacking	Insecure interfaces and APIs	1.Identity and Access Management Guidance [10] 2.Dynamic credential[11]
2	SAAS PASS IAAS	Data leakage	1.Data-related vulnerabilities 2.Vulnerabilities in Virtual Machines 3.Vulnerabilities in Virtual Machine Images 4.Vulnerabilities in Virtual Networks	FRS techniques [12] Digital Signatures [13] Encryption [14] Homomorphic encryption [14]
3	SAAS PASS IAAS	Denial of Service	1.Insecure interfaces and APIs 2.Unlimited allocation of resources	Cloud providers can force policies to offer limited computational Resources 1.Filtering tree 2. Trace-back and filter system
4	SAAS	Customer-data Manipulation	Insecure interfaces and APIs	Web application scanners [15]
5	IAAS	VM escape	Vulnerabilities in Hypervisors	HyperSafe [16] TCCP (Trusted Cloud Computing

				Platform) [17]
6	IAAS	Malicious VM creation	Vulnerabilities in Virtual Machine Images	Mirage [18]
7	IAAS	Insecure VM migration	Vulnerabilities in Virtual Machines	PALM [19] VNSS [20]
8	IAAS	Sniffing/Spoofing virtual networks	Vulnerabilities in Virtual Networks	Virtual network framework based on Xen network modes: "bridged" and "routed" [21]

4.1 Threats: Account or service hijacking

Due to Insecure interfaces and APIs vulnerability in SAAS, PAAS, IAAS Account or service Hijacking Attack is made, the possible solutions for which are 1.Identity and Access Management Guidance [10] 2.Dynamic credential [11] as detailed below.

A record theft can be performed by various ways, for example, social building and powerless accreditations. In the event that an aggressor accesses a client's qualification, he can perform noxious exercises, for example, get to delicate information, control information, and divert any exchange [8].

4.1.1 Vulnerabilities: Insecure interfaces and APIs

Cloud providers present services that can be used through APIs (. The security of the cloud depends upon the security of these interfaces [8]. Some problems are:

1. powerless accreditations
2. Insufficient authorization checks
3. Insufficient input-data validation

4.1.2 Countermeasures for Account or service hijacking

Below are the solutions of account or service hijacking

1. Identity and Access Management Guidance: Cloud Security Alliance (CSA) is a non-benefit association that advances the utilization of best practices so as to give security in cloud situations. CSA has given an Identity and Access Management Guidance [10] which gives a rundown of prescribed best rehearsed to guarantee personalities and secure access the executives. This report incorporates centralized directory, access management, identity management, role-based access control, user access certifications, privileged user and access management, separation of duties, and identity and access reporting.

2.Dynamic Credentials: [11] presents a algorithm to make dynamic accreditations for mobile cloud computing frameworks. The dynamic qualification changes its worth once a client changes its area or when he has traded a specific number of information parcels.

4.2. Threats: Data leakage

Due to Data-related vulnerabilities in Virtual Machines, Vulnerabilities in Virtual Machine Images Vulnerabilities in Virtual Networks on SAAS, PAAS, IAAS data leakage attack is made, the possible solutions for which are FRS techniques [12],Digital Signatures [13],Encryption [14],Homomorphic encryption [14]

Data leakage happens when the data gets into the wrong hands while it is being transferred, stored, audited or processed [8].

4.2.1Vulnerabilities

Here are some vulnerabilities in Data leakage

1. Unlimited allocation of resources-

Inaccurate modeling of resource usage can lead to overbooking or over-provisioning [2].

2. Data-related vulnerabilities:

Here in this Type of Vulnerabilities Data can be collocated with unknown owners, Incomplete data

cannot be completely removed, Data backup done by untrusted user

Data deduplication: a technique that stores only a copy of redundant data which may be not secured

3. Vulnerabilities in Virtual Machine Images

A. Uncontrolled placement of VM images in public repositories [7]

4. Vulnerabilities in Virtual Machines

1. Possible covert channels in the collocation of VMs
2. Unrestricted allocation and deallocation of resources with VMs [2]
3. Uncontrolled Migration - VMs can be migrated from one server to another server due to fault tolerance, load balance, or hardware maintenance
4. Uncontrolled snapshots – VMs can be copied in order to provide flexibility, which may lead to data leakage
5. Uncontrolled rollback could lead to reset vulnerabilities

4.2.2 Countermeasures for: Data Leakage

There are mainly three solutions for Data leakage attack which are listed below.

1. Fragmentation-redundancy-scattering (FRS) technique [12]: This method means to give interruption resilience and, in result, secure storage. This method comprises in first separating delicate information into irrelevant parts, so any piece doesn't have any noteworthy data independent from anyone else. At that point, pieces are dissipated in a repetitive style crosswise over various locales of the disseminated framework.

2. Digital Signatures: [13] proposes to verify information utilizing computerized signature with RSA calculation while information is being moved over the Internet. They guaranteed that RSA is the most conspicuous calculation, and it very well may be utilized to secure information in cloud situations

3. Homomorphic encryption: The three fundamental activities for cloud information are move, store, and procedure. Encryption strategies can be utilized to verify information while it is being moved all through the cloud or put away in the supplier's premises. Cloud suppliers need to unscramble figure information so as to process it, which raises protection concerns. In [14], they propose a technique dependent on the utilization of completely homomorphic encryption to the security of mists. Completely homomorphic encryption permits performing subjective calculation on figure writings without being unscrambled. Current homomorphic encryption plans bolster predetermined number of homomorphic tasks, for example, expansion and duplication.

4.3 Threat: Denial of Service

Due to the Vulnerabilities in insecure interface and APIs and unlimited resource allocation denial of service attack is made the possible solution for which is filtering tree and trace back filtering system

It is possible that a malicious user will take all the possible resources. Thus, the system cannot satisfy any request from other legitimate users due to resources being unavailable.

4.3.1 Vulnerabilities

Here are some Vulnerabilities of denial service

1. Insecure interfaces and APIs

Cloud providers offer services that can be accessed through APIs (SOAP, XML/JSON). The security of the cloud depends upon the security of these interfaces [16]. Some problems are:

A. Weak credential

B. Insufficient authorization checks

C. Insufficient input-data validation

2. Unlimited allocation of resources- Inaccurate modeling of resource usage can lead to overbooking or over-provisioning

4.3.2 Countermeasures for Threat: Denial of Service (Denial-of-Service Prevention Mechanisms)

There are two possible solutions which are filtering tree and trace back filtering system which are listed below

1. **A filtering tree** that goes about as an assistance intermediary inside a Service Oriented Architecture (SOA) model is displayed. The creators inspect the uncertainties in institutionalized cloud APIs and how these can be misused in provisioning, the board and checking of administrations. They propose adding a mark reference component to each SOAP solicitation to guarantee that it originates from a genuine source. Twofold marks are created utilizing hashed qualities of each SOAP envelope, for example, the quantity of youngster or header components. The customer IP address is kept in the message header alongside a riddle that is put away as a component of the WSDL record. The proposed framework needs to examine every bundle exclusively, which can prompt a bottleneck in DDoS circumstances.

2. **A trace-back and filter system is proposed** to protect the cloud from DDoS attacks. SOA trace-back is used by adding a tag to SOA packets to record the route taken. This system cannot identify the source of attack, because the tag is only added to the packet once it is relatively close to the server. The tests used in the paper do not consider spoofed IP addresses or the fact that an attacker is likely to make use of zombie machines

4.4 Threat: Customer Data Manipulation

Due to Data-related vulnerabilities in insecure interfaces and APIs on SAAS Customer Data Manipulation attack is made, the possible solutions for which are Web application scanners [15]

Users attack web applications by manipulating data sent from their application component to the server's application. For example, SQL injection, command injection, insecure direct object references, and cross-site scripting

4.4.1 Vulnerabilities

Vulnerabilities of Customer Data manipulation is Insecure interfaces and APIs

1 Insecure interfaces and APIs

Cloud providers present services that can be used through APIs

4.4.2 Countermeasures for Threat: Customer Data Manipulation

Here is the possible solution for customer data manipulation

1. **Web application scanners**: Web applications can be an easy target because they are exposed to the public including potential attackers. Web application scanners [15] is a program which scans web applications through the web front-end in order to identify security vulnerabilities. There are also other web application security tools such as web application firewall. Web application firewall routes all web traffic through the web application firewall which inspects specific threats.

4.5. Threat: VM Escape

Due to Data-related vulnerabilities in Vulnerabilities in Hypervisors on IAAS VM Escape attack is made, the possible solutions for which are Hyper Safe [16] TCCP (Trusted Cloud Computing Platform) [17]

It is designed to exploit the hypervisor in order to take control of the underlying infrastructure [1]

4.5.1 Vulnerabilities in Hypervisors

1. Complex hypervisor code [1]

2. Flexible configuration of VMs or hypervisors to meet organization needs can be exploited

4.5.2. Countermeasures for Threat: VM Escape

1. Hyper Safe [16]: It is a methodology that gives hypervisor control-stream trustworthiness. Hyper's Safe will probably ensure type hypervisors utilizing two procedures: non-by acceptable memory lockdown which shields compose shielded memory pages from being adjusted, and limited pointed ordering that changes over control information into pointer records. So as to assess the adequacy of this methodology, they have led four kinds of assaults, for example, adjust the hypervisor code, execute the infused code, alter the page table, and alter from an arrival table. They inferred that Hyper Safe effectively anticipated every one of these assaults, and that the presentation overhead is low.

2.**Trusted Cloud Computing Platform**: TCCP [17] enables suppliers to offer shut box execution conditions, and enables clients to decide whether nature is secure before propelling their VMs. The TCCP includes two basic components: a trusted virtual machine monitors (TVMM) and a trusted coordinator (TC). The TC deals with a lot of confided in hubs that run TVMMs, and it is kept up yet a confided in outsider. The TC partakes during the time

spent propelling or moving a VM, which confirms that a VM is running in a confided in stage.

4.6. Threat: Malicious Virtual Machine Creation

Due to Virtual Machine Images vulnerability on IAAS Malicious Virtual Machine Creation attack is made, the possible solutions for which is Mirage [18]

An attacker who creates a legitimate account can make a VM image containing malicious code such as a Trojan horse and store it in the provider storehouse [1].

4.6.1 Vulnerabilities in Virtual Machine Images

1.Uncontrolled placement of VM images in public repositories [7]

2. VM images are not able to be patched since they are dormant artifacts

4.6.2. Countermeasures for Malicious Virtual Machine Creation

1.Mirage: In [18], the authors suggest a virtual machine image management system in a cloud computing atmosphere. This technique includes security features: access control framework, image filters, a provenance tracking, and repository maintenance services.

4.7. Threats: Insecure Virtual Machine Migration

Due to vulnerabilities in Virtual Machine on IAAS Insecure Virtual Machine Migration attack is made, the

possible solutions for which are Protection Aegis for Live Migration of VMs [19] and VNSS [20]

Live migration of virtual machines exposes the contents of the VM state files to the network. An attacker can do the

following actions:

1. Access data illegally during migration [2]

2. Transfer a VM to an untrusted host [2] Create and migrate several VM causing disruptions or DoS

4.7.1 Vulnerabilities in Virtual Machines

Uncontrolled snapshots – VMs can be copied in order to provide flexibility, which may lead to data leakage

4.7.2 Countermeasures for Threat: Insecure Virtual Machine Migration

Here are the possible solutions for Insecure Virtual Machine Migration

1. Protection Aegis for Live Migration of VMs (PALM): [19] proposes a secure live migration framework that jam uprightness and security insurance during and after movement. The model of the framework was executed dependent on Xen and GNU Linux, and the consequences of the assessment demonstrated that this plan just includes slight personal time and relocation time because of encryption and unscrambling.

2.VNSS: [20] proposes a security framework that customizes security policies for each virtual machine, and it

provides continuous protection thorough virtual machine live migration. They implemented a prototype system

based on Xen hypervisors using stateful firewall technologies and user space tools such as ip tables, xm commands

program and conntrack-tools..

4.8. Threat: Sniffing/Spoofing virtual networks

Due to vulnerabilities in Virtual Networks on IAAS Sniffing/Spoofing virtual networks attack is made, the possible solutions for which is Virtual Network Security [21]

A malicious VM can listen to the virtual network or even use ARP spoofing to redirect packets from/to other VMs.

4.8.1 Vulnerabilities in Virtual Networks

Sharing of virtual bridges by several virtual machines [21]

4.8.2 Countermeasures for Threats: Sniffing/Spoofing virtual networks

Here is the list of possible solutions for Sniffing/Spoofing virtual networks

1. Virtual Network Security: Wu and et al [21] presents a virtual network framework that secures the communication among virtual machines. This system depends on Xen which offers two design modes for virtual systems: "spanned and directed" The virtual system model is made out of three layers: steering layers, firewall, and shared systems, which can counteract VMs from sniffing and caricaturing. An assessment of this methodology was not performed when this production was distributed.

VII. CONCLUSION

As described in the paper, however, there are exciting advantages and benefits of cloud computing, there are yet several security problems that reduce his adoption. Traditional methods of security cannot solve all kind of cloud environment security problems because it is a complex architecture composed of a combination of different technologies. Therefore, thinking about new security methods and policies to include in cloud architecture is a necessity today.

References

[1] Mazhar Ali , Samee U. Khan , Athanasios V. Vasilakos b , 2015, "Security in cloud computing: Opportunities and challenges" Elsevier transaction on Information Science 305(2015) ,pp.357-383

[2] Keiko Hashizume, David G Rosado, Eduardo Fernández-Medina and Eduardo B Fernandez, "An analysis of security issues for cloud computing" Springer Journal of Internet Services and Applications 2013, 4:5

[3]. Ju J, Wang Y, Fu J, Wu J, Lin Z (2010) Research on Key Technology in SaaS. In: International Conference on Intelligent Computing and Cognitive Informatics (ICICCI), Hangzhou, China. IEEE Computer Society, Washington, DC, USA, pp 384–387

[4].OWASP (2010) The Ten most critical Web application Security risks. Available: https://www.owasp.org/index.php/Category: OWASP Top Ten Project

[5]. Cloud Security Alliance (2012) Security guidance for critical areas of Mobile Computing. Available:https://downloads.cloudsecurityalliance.org/initiatives/mobile/Mobile_Guidance_v1.pdf

[6]. Ertaul L, Singhal S, Gökay S (2010) Security challenges in Cloud Computing. In: Proceedings of the 2010 International conference on Security and Management SAM'10. CSREA Press, Las Vegas, US, pp 36–42

[7]. Morsy MA, Grundy J, Müller I (2010) An analysis of the Cloud Computing Security problem. In: Proceedings of APSEC 2010 Cloud Workshop. APSEC, Sydney, Australia

[8]. Cloud Security Alliance (2010) Top Threats to Cloud Computing V1.0. Available: https://cloudsecurityalliance.org/research/top-threats

[9]. Hashizume K, Yoshioka N, Fernandez EB (2013) Three misuse patterns for Cloud Computing. In: Rosado DG, Mellado D, Fernandez-Medina E, Piattini M (ed) Security engineering for Cloud Computing: approaches and Tools. IGI Global, Pennsylvania, United States, pp 36-53

[10]. Cloud Security Alliance (2012) SecaaS implementation guidance, category 1: identity and Access managament.Available:https://downloads.cloudsecurityalliance.org/initiatives/secaas/SecaaS_Cat_1_IAM_Implementation_Guidance.pdf

[11]. Xiao S, Gong W (2010) Mobility Can help: protect user identity with dynamic credential. In: Eleventh International conference on Mobile data Management (MDM). IEEE Computer Society, Washington, DC, USA, pp 378–380

[12]. Wylie J, Bakkaloglu M, Pandurangan V, Bigrigg M, Oguz S, Tew K, Williams C, Ganger G, Khosla P (2001) Selecting the right data distribution scheme for a survivable Storage system. CMU-CS-01-120, Pittsburgh, PA

[13]. Somani U, Lakhani K, Mundra M (2010) Implementing digital signature with RSA encryption algorithm to enhance the data Security of Cloud in Cloud Computing. In: 1st International conference on parallel distributed and grid Computing (PDGC). IEEE Computer Society Washington, DC, USA, pp 211–216

[14] Jun Tang, Yong Cui, Qi Li, Kui Ren, Jiangchuan Liu, and Rajkumar Buyya. 2016."Ensuring security and privacy preservation for cloud data services".ACM Computing. Survey. 49, 1, Article 13 (June 2016), 39 pages.

[15]. Fong E, Okun V (2007) Web application scanners: definitions and functions. In: Proceedings of the 40th annual Hawaii International conference on system sciences. IEEE Computer Society, Washington, DC, USA

[16]. Wang Z, Jiang X (2010) HyperSafe: a lightweight approach to provide lifetime hypervisor control-flow integrity. In: Proceedings of the IEEE symposium on Security and privacy. IEEE Computer Society, Washington, DC, USA, pp 380–395

[17]. Santos N, Gummadi KP, Rodrigues R (2009) Towards Trusted Cloud Computing. In: Proceedings of the 2009 conference on Hot topics in cloud computing, San Diego, California. USENIX Association Berkeley, CA, USA

[18]. Wei J, Zhang X, Ammons G, Bala V, Ning P (2009) Managing Security of virtual machine images in a Cloud environment. In: Proceedings of the 2009 ACM workshop on Cloud Computing Security. ACM New York, NY, USA, pp 91–96

[19]. Zhang F, Huang Y, Wang H, Chen H, Zang B (2008) PALM: Security Preserving VM Live Migration for Systems with VMM-enforced Protection. In: Trusted Infrastructure Technologies Conference, 2008. APTC'08,

Third Asia- Pacific. IEEE Computer Society, Washington, DC, USA, pp 9–18

[20].Xiaopeng G, Sumei W, Xianqin C (2010) VNSS: a Network Security sandbox for virtual Computing

environment. In: IEEE youth conference on information Computing and telecommunications (YC-ICT). IEEE Computer Society, Washington DC, USA, pp 395–39

Signature Based Data Sharing Over the Cloud

Priti Rumao[1] , Mamta Meena[2] and Samidha Kurle[3]
[1,2,3] Computer Engineering Department,
Atharva College of Engineering, Malad West,
Mumbai- 400095, India

Abstract. **Cloud computing, which is subscription based service, mainly focuses on data sharing from any place, anywhere. So when data sharing amongst multiple participants come into the picture, it faces multiple issues associated with data integrity, efficiency and privacy of data owner. The ring signature is inspiring concept to model authenticate and anonymous data sharing system. The basic functionality of ring signature is to furnish signer's privacy and us-er's anonymity. Still the process of certificate verification remains a bottleneck which doesn't allow this system to be more scalable. Hence, the ID based ring signature can be used as alter-native solution for this. The concept of forward security has been added further to make ID based ring signature more secure, because even if a key has been compromised, all previously generated signatures remain valid. In this signature based data sharing over cloud system, for-ward secure ID-based ring signature is further enhanced by adding weil pairing: It keeps even shorter keys secure, it also requires less processing power and time. This property is essential to build large scale data sharing system. So here efficient and tangible application of this system has been provided.**

Keywords: Authentication, data sharing, cloud computing, forward security

I. INTRODUCTION

1.1 Overview

Recent trends in technology, generates huge amount of data. Cloud computing is flourishing technology with many characteristics, amongst which data storage and data sharing provides number of benefits to both i.e. individuals as well as society. Now-a-days cloud is mostly used for data storage and data sharing. Through cloud computing, huge number of participants gains large number of data. But many partic-ipants may lead to compromising data integrity, efficiency and privacy. So to over-come issues related to above mentioned key features, Rivest, Shamir and Tauman proposed a concept of ring signature[2]. But in case of ring signatures, need of certifi-cates validity verification and management was important, which leads to spending much more cost on it. The concept of Identity-Based (Id-Based) cryptosystem [3] introduced by Shamir to overcome certificates verification and management issues. In this, for each user the public key can be easily computed from data related to this user's publicly known identity. From its master secret for users, a private key genera-tor (PKG) then generates private keysfor users. This property avoids the need of certificates and accompanies an innate public key to each user within the system.Id-Based system, eliminates the need of certificates verification, hence saves communi-cation and computational cost.When combining ring signature with Id-based system, the outcome has much more advantages than certificates based ring signature system. In case of traditional certificate-based ring signature, verifier has to validate each and every certificate of the members (users) present in the ring. Once certificate validation is done, then verifier will verify message and user's signature. For Id-based ring sig-nature, entire certificate validation process is eliminated. Here verifier just has to verify identity of the user and message-signature pair of user. Since certificate valida-tion process has been eliminated, it saves hugely time and computation cost. This system is most beneficial to organizations having huge data and many more users. The first ever ID-based ring signature scheme was proposed in 2002 by Zhang and Kim [4]. Other proposed existing ID-based ring signature schemes include [5], [6], [7], [8], [9], [10], [11], [12].

For data sharing, more users can be added to ring. More members are added to provide more protection. But by increasing number of users, probability of key expo-sure also increases. So if one private key of the signer is compromised then all gener-ated signatures of that particular signer becomes valueless i.e. future signatures are no validated and previously issued signatures cannot be trusted. Hence when a key leak-age is acknowledged, key revocation mechanisms must be invoked immediately to prevent the generation of any signature using the compromised secret key. Still, this does not solve the problem of past signatures which has been forged.

Hence, to conserve the validity of past signatures even if current signature is com-promised, new concept of forward secure signature has been introduced by Anderson [13], and the solutions were developed by Bellare and Miner [14].The basic idea in-troduced by Bellare and Miner is: Total validity time (T) of public key is divided into smaller time

slots. Even if during current time slot a key is compromised, then for adversary it is difficult to find out previous or next keys. Huang et al. [1] combine Id-based ring signature with forward security to generate robust solution. In this, in case of ring signature, importance of forward security is highlighted and robust algorithm has been designed. It is first solid design of forward secure Id- based ring signature and under the standard RSA assumption, its security has been tested (proved) in ran-dom oracle model. Still it is always possible to reduce cost with respect to time con-sumption.

Andre Weil introduced the notion of weil pairing in 1940. This concept was mainly designed to attack elliptic curve cryptography (ECC) system [15].Sakai et al. [16] showed that weil pairing can be used for key exchange. Boneh et. al. [17] developed identity-based encryption scheme which has chosen cipher text security in the random oracle model, in which encryption is done by modified Elgamal and decryption is done by modified weil pairing computation.

So, to summarize, signature based data sharing over cloud makes use of advantages of all three techniques, namely, Id based ring signature, forward security and weil pairing to develop robust solution.

1.2 Contribution

In this paper, a new notion called signature based data sharing over cloud has been proposed, which proves its importance for building cost effective authentic and anon-ymous system to mainly save time consumption for data sharing system.

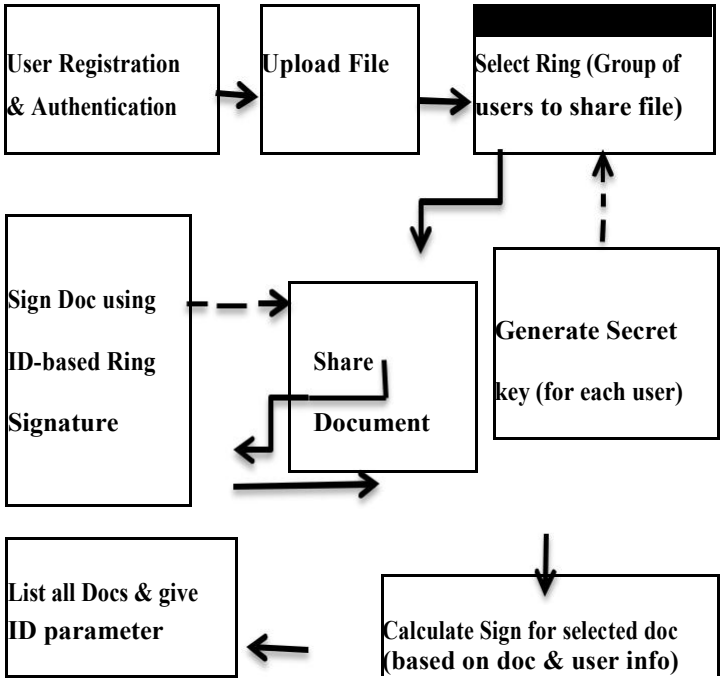
The system presents concrete solution for Id based ring signature with forward security and weil pairing to make it more robust.

Practical implementation of the system is possible:

- Since it uses Id based ring signature scheme, so cost of certificate validation has been removed.
- Secret key size is small. Key update process is easy but generates needed result.
- Use of weil pairing further reduces time consumption but enhances security process.

II. PROPOSED SYSTEM

Signature based data sharing over cloud architecture proposed an efficient data sharing system in cloud environment which supports data authentication and anonym-ity of signer. The workflow of proposed system is shown in figure 2.1. In this, basi-cally first data owner registers and authenticates users into the system. Currently sys-tem is single owner multi users system. Once registration and authentication is done, owner uploads file to the system. Once file has been uploaded to the system, owner selects group of users with whom, document (data) has to be shared and generate their secret key for the ring. While sharing the document, using id-based ring signature scheme, the document has been signed and shared. User will be able to see all the shared documents. While downloading the document, previously calculated signature and signature of selected document has been verified. Once signature verification is successful document will be downloaded.



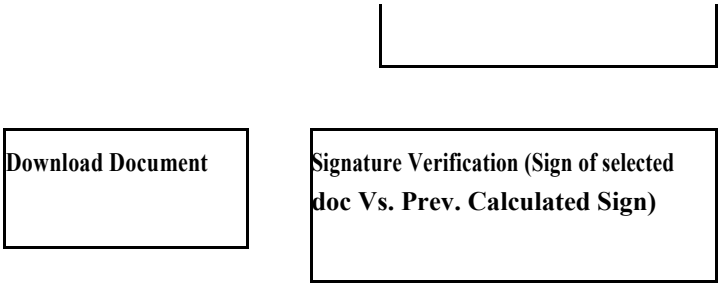


Fig 2.1: Proposed system of Signature based data sharing over cloud

The following steps show the workflow of the proposed system:

Step 1: User registers into the system: At this stage concept of forward security has been added. Here for public id i.e. Username total validity period of a year is as-signed. That entire year has been divided into time slots, during which secret key i.e. password of user will remain valid. One day prior to end of current time slot, system will request for change of password. At the end of the time slot unless and until pass-word has been changed, system won't allow users' login. At the end of the public key validity, user simply has to register and authenticate into the system. (Refer figure 2.2).

Fig 2.2: User Registration

Step 2: Data owner uploads file to be shared into the system (Refer figure 2.3).

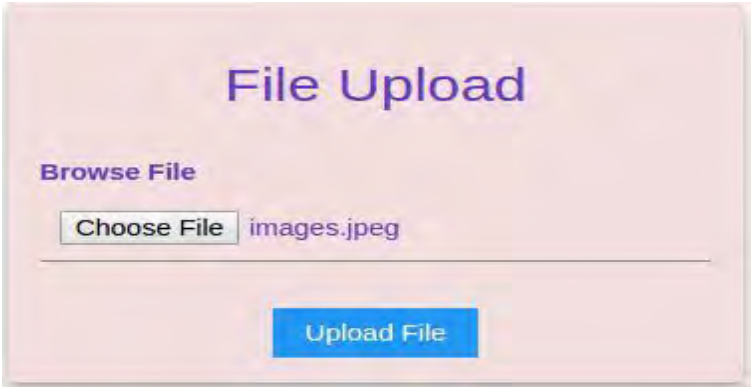


Fig 2.3: Upload File

Step 3: Data owner selects group of users (ring) to share selected file: Here using the concept of weil pairing ring (group) signature has been generated and to download the document, this signature with document and user's certain information has been send to selected users (Refer figure 2.4).

Fig 2.4: Users Selection to share data

Step 4: Data owner share documents amongst selected group of users: When doc-ument is shared, that time, based on all the keys of selected users and document in-formation, one signature for that document has been generated (Refer figure 2.5).7

Fig 2.5: File sharing done.

Step 5: All the shared document has been shared with all the users in list format (Refer figure 2.6).

Fig 2.6: Shared data

Step 6: To download the document, first based on the document and user infor-mation, signature has been generated and then signature of currently selected document and previously generated signature associated with that document has been compared.

Step 7: Once both the signatures are matched, then document will be downloaded for that particular user: Here only if signature is matched, then only user will be able to download the document or else user is not allowed to access that particular document (Refer figure 2.7).

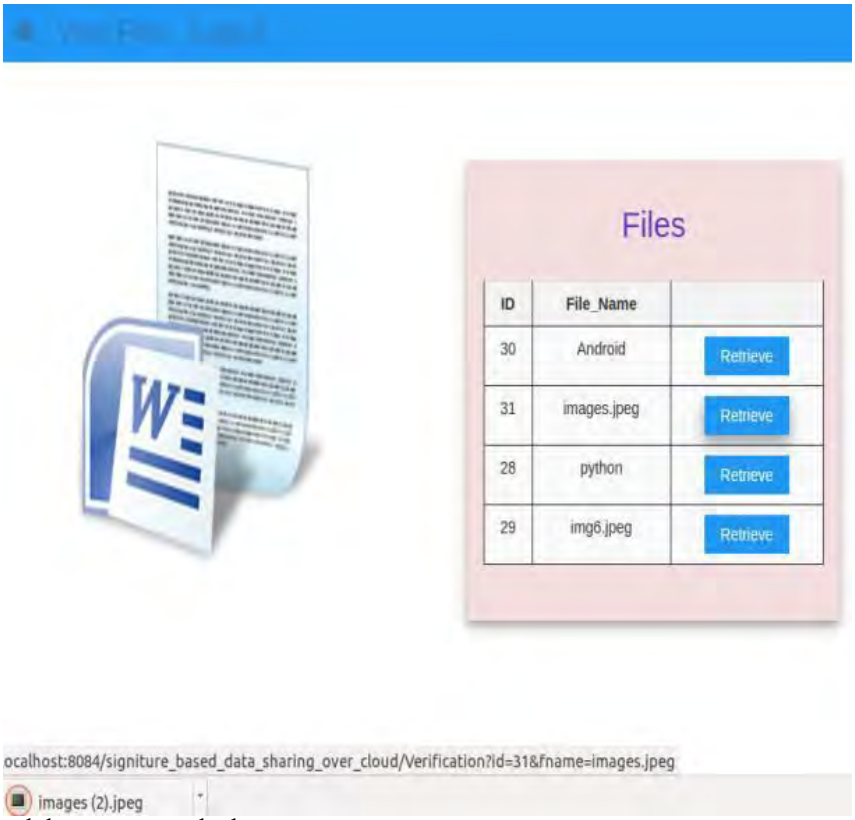


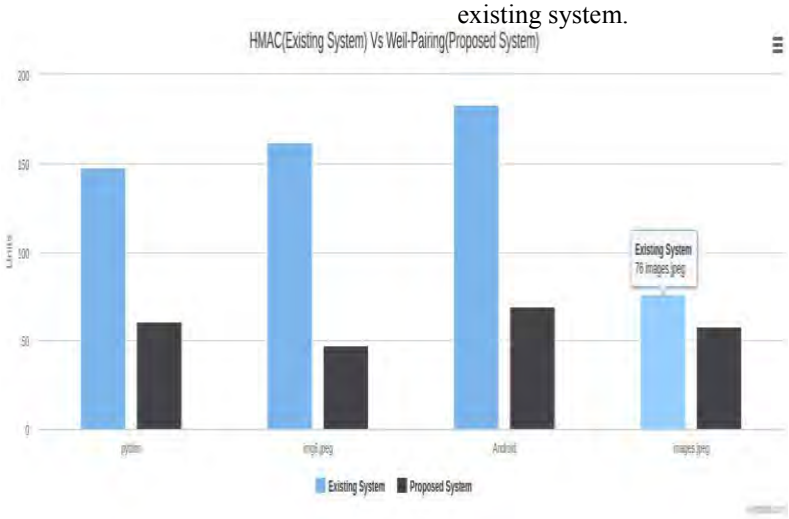
Fig 2.7: Download data once matched.

III. COMPARATIVE ANALYSIS

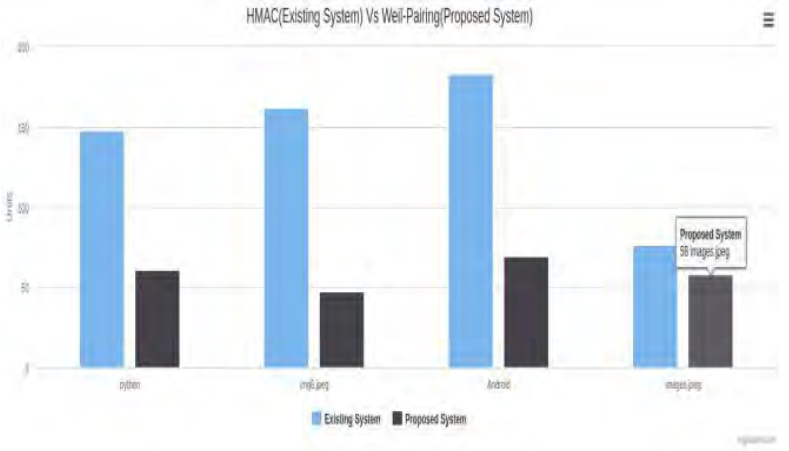
In this section, analysis of proposed system with respect to exiting system has been done. This analysis has been done on the basis of execution time parameter. The existing forward secure id- based ring signature system [1] is based on two different schemes namely, forward security and id- based ring

signature, whereas proposed system is combination of weil pairing, forward security and id- based ring signature. Proposed system supports and implements the concept of pairing, whereas existing system avoids pairing concept altogether.

The results, comparing existing system and proposed system has been shown in fig 3.1 (a) & (b). It can be seen that document retrieving time for existing system is high when compared with proposed system. So is can be said that proposed system provide cost efficient solution than the existing system.



(a)



(b)

Fig. 3.1 (a) (b): Graphical representation for time required to retrieve documents between existing and proposed system

IV. CONCLUSION

In this paper, to fulfill the ever growing need of data sharing, a new scheme called signature based data sharing over cloud has been proposed. It combines id- based ring signature with forward security and weil pairing. This scheme provides anonymity, authentication as well as reduce execution time parameter. With the help of weil pairing, proposed system becomes more efficient. This scheme will be very useful in many practical applications where authentication and user's anonymity plays an important role. The proposed scheme currently focuses on re-ducing execution time parameter. Providing more security features and still reducing cost re-garding time and space parameter will be the open problem and future research work.

References

1. X. Huang, J. Liu, S. Tang, Y. Xiang, K. Liang, L. Xu, Andj. Zhou, "Cost-Effective Au-thentic And Anonymous Data Sharing With Forward Security," *Computers, Ieee Transac-tions On*, 2015, Vol: 64 No: 6 Year 2015.
2. A. Shamir R. Rivest And Y. Tauman,"How To Leak Secret", In Asiacrypt'01 Lncs 2248 (2001), 552,565.
3. A. Shamir,"Identity-Based Cryptosystems And Signature Schemes". In Crypto 1984,
   Volume 196 Of Lecture Notes In Computer Science, Pages 47–53. Springer, 1984.
4. F. Zhang And K. Kim,"Id-Based Blind Signature And Ring Signature From Pairings". In Asiacrypt 2002, Volume 2501 Of Lecture Notes In Computer Science, Pages 533–547. Springer, 2002.
5. H.-Y. Chien."Highly Efficient Id-Based Ring Signature From Pairings".In Apscc, Pages
   829–834, 2008.
6. A. K. Awasthi And S. Lal. "Id-Based Ring Signature And Proxy Ring Signature Schemes
   From Bilinear Pairings". Corr, Abs/Cs/0504097, 2005.
7. J. Zhang. "An Efficient Identity-Based Ring Signature Scheme And Its Extension". In Ic-csa (2), Volume 4706 Of Lecture Notes In Computer Science, Pages 63–74. Springer, 2007.
8. S. S. M. Chow, S.-M. Yiu, And L. C. K. Hui. Efficient Identity Based Ring Signature. In Acns 2005, Volume 3531 Of Lecture Notes In Computer Science, Pages 499–512. Spring-er, 2005.
9. S. S. Chow, R. W. Lui, L. C. Hui, And S. Yiu. Identity Based Ring Signature: Why, How And What Next. In D. Chadwick And G. Zhao, Editors, Europki, Volume 3545 Of Lecture Notes In Computer Science, Pages 144–161. Springer, 2005.
10. L. Nguyen. Accumulators From Bilinear Pairings And Applications. In A. J. Menezes, Editor, Ct-Rsa 2005, Volume 3376 Of Lecture Notes In Computer Science, Pages 275–
292. Springer, 2005.
11. J. Herranz,"Identity-Based Ring Signatures From Rsa". Theor. Comput. Sci., 389(1-
    2):100–117, 2007.
12. P. P. Tsang, M. H. Au, J. K. Liu, W. Susilo, And D. S. Wong,"A Suite Of Non-Pairing Id-Based Threshold Ring Signature Schemes With Different Levels Of Anonymity" (Extend-ed Abstract). In Provsec, Volume 6402 Of Lecture Notes In Computer Science, Pages 166–183. Springer, 2010.
13. R. Anderson,"Two Remarks On Public-Key Cryptology. Manuscript", Sep. 2000. Rele-vant Material Presented By The Author In An Invited Lecture At The Fourth Acm Confer-ence On Computer And Communications Security, 1997.
14. M. Bellare And S. Miner,"A Forward-Secure Digital Signature Scheme". In Crypto'99,
    Volume 1666 Of Lecture Notes In Computer Science, Pages 431–448. Springer-Verlag, 1999.
15. G. Frey, M. Muller, H. Ruck, "The Tate Pairing And The Discrete Logarithm Applied To
    Elliptic Curve Cryptosystems", Ieee Tran. On Info. Th., Vol. 45, Pp. 1717-1718, 1999.
16. R. Sakai, K. Ohgishi, And M. Kasahara,"Cryptosystems Based On Pairings," In Proceed-ings Of Symposium On Cryptography And Information Security, Japan, 2000.
17. Dan Boneh, Matthew Franklin, "Identity-Based Encryption From The Weil Pairing", Ap-pears In Siam J. Of Computing, Vol. 32, No. 3, Pp. 586-615, 2003.

Driver Activity Monitoring using MobileNets

Garima Tripathi, Deval Srivastav, Saim Shaikh, Priyank Shah
Dept. of Information Technology
Fr. Conceicao Rodrigues College of Engineering,
Mumbai,Maharashtra,India
garima@frcrce.ac.in

Abstract: **This paper presents a method to monitor the driver's activity and continuously look for red flags such as distracted driving, overuse of mobile phones while driving, drowsiness, sleeping. This is achieved by using a camera based system and the mobilenet neural network which has been trained on our self- made dataset.**

Keywords: Computer Vision, Machine Learning, Activity Monitoring, MobileNets, Image Processing.

I. Introduction

In our current world as driving technology continues to grow the driving effort required decreases. Our drivers become more and more careless resulting in loss of life in many circumstances. The proposed system aims to solve this problem by developing a system to monitor driver's activity and warn him when necessary. The method involves deploying a neural network trained on categories such as talking or texting on the phone talking to co passengers, operating the radio and drinking water. In our paper we have extensively tested performance of different neural networks[1] such as resnet-50[2] , inception[3] and also mobile nets[7]. Throughout the development our focus has been on how close we can get to real life conditions so our network will receive the images from an IR camera allowing our system to perform during night time. Our system can be easily fitted to any existing vehicle very easily and will be intuitive to use. The system will also be equipped with sensors to detect, rash driving and will consist of security features such as fencing, fingerprint authentication to prevent thieving of the vehicle.

In this domain lots of research has been done it can be concluded that the most popular vision methods include detecting driver inattention using head pose eye gaze estimation or simply checking eye closure rate. Measures such as EEG, electrocardiogram, etc. Our paper will discuss these methods and other techniques that have been used in the next section.

The paper has been organized in the following manner, the next section will discuss related work in the given domain in depth after which we will present our method of approaching this condition and finally we will discuss all the results we have obtained from the proposed system.

II. Literature Survey

According to the research conducted by us it can be concluded that the most popular methods to solve this problem involve either driver biological measures, driver physical measures, driving performance measures or some kind of a hybrid measure[3].

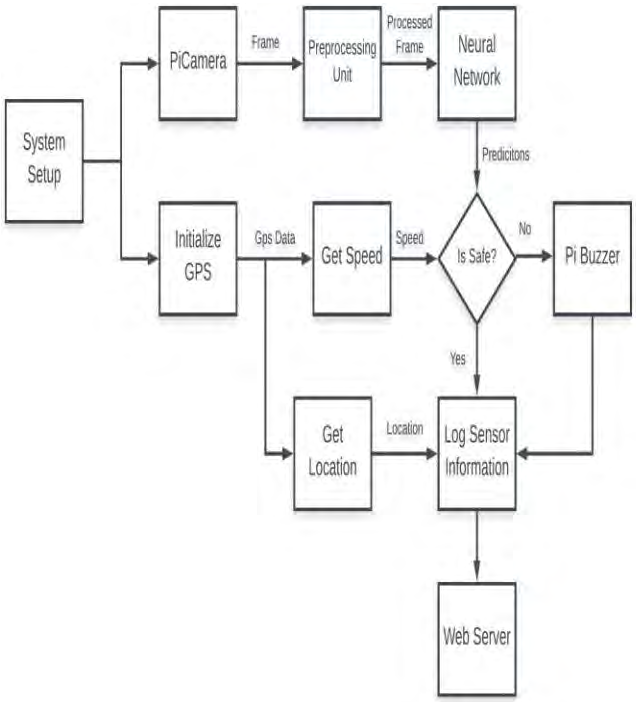
Driver biological measures include biological signals like EEG, electrocardiogram (ECG), electro-oculography (EOG). These signals are collected through electrodes in contact with the skin and then analysed for fatigue and drowsiness. Physical measures involve eye closure detection and blink frequency, face position, driver gaze to detect inattention.

Driver performance measures involve various measures such as steering angle and other driving criterion. Most research that has been done in the related field has been focused on detecting driver inattention using eye gaze tracking and head pose estimation. These methods rely only on head and eye movement to detect inattention whereas in real life a driver can be distracted doing various tasks that cannot be detected by head movement alone. It has been observed that current driver monitoring systems employ statistical machine learning methods to detect driver distraction and work on a limited dataset .Some Research has been done on applying deep learning technologies to solve this problem but such systems cannot be deployed in a vehicle in a cost effective manner nor do they work in night time conditions[4][5].

III. Proposed Method

Figure 1 describes the entire workflow of the project right from the hardware setup to the user interface. The Pi camera is mounted on an appropriate position in the dashboard of the vehicle. It is then connected to the camera port on the Raspberry Pi. The SM808 GSM +GPS module is connected to the Raspberry Pi via USB TO RS232 serial port. The GPS antenna is connected to the module and placed outside the

vehicle with a clear view of the sky Connect the Raspberry Pi to a portable power supply via the micro USB port. The Pi camera records footage and send the frames to the Preprocessing Unit .The footage is recorded at a resolution of 640*480 at 25 frames per second. For the night time footage the Pi Camera is aided by 2 IR bulbs which help provide clear frames even in pitch black conditions. The frames received from the Pi Camera are processed .The frames are first flipped



and then are resized to the size of 224*224 as the network takes an input of 224*224 images.

Fig.1: Implementation flowchart

TABLE 1
DATASET STATISTICS

Sr.no	Class	Count(day +night)
1.	Safe Driving	5000
2.	Talking on phone	5000
3.	Texting on phone	5000
4.	Drinking	5000
5.	Sleeping	5000
7.	Yawning	5000

A. Mobilenet

Mobilenet[7] is a neural network that was developed by google to perform on low powered devices lacking graphical GPUs that are known to accelerate neural network performance. MobileNets are small, low-latency, low-power models parameterized to meet the resource constraints of a variety of use cases, one of those use case devices can also be deploying it on a raspberry pi which we intend to do.

A standard convolution both filters and combines inputs into a new set of outputs in one step, but in a case of mobilenets it first uses depth wise convolution that applies a single filter to each input channel. The point wise convolution then applies a 1X1 convolution to combine the outputs the depth wise convolution. The depth wise separable convolution splits this into two layers, a separate layer for filtering and a separate layer for combining. This factorization has the effect of drastically reducing computation and model size, this modification allows mobilenet to be faster than its other counterparts.

For our application we have employed a mobilenet v2 , its the second iteration of mobilenets and now it along with the depth wise separable blocks it also uses bottleneck residual layers and also adds a 1 x 1 expansion whose purpose is to expand the number of channels in the data before it goes to the next block. In the proposed system we have used mobilenet v2 as its much more better than its older version, mobilenet was trained on a self made dataset of driver's performing distracted activities.

Fig. 2. Dataset samples for night

Fig. 3. Dataset samples for day

In the above pictures we can see the dataset samples from night and from day. The images are in order of Sleeping, Talking on the phone, Drinking, Texting, Yawning,

B. Training

This section will discuss how we trained our model on the above dataset. We have used transfer learning to

train our neural network as we have used model of mobilenet v2 and only trained the last few layers to get best results and quicker training times.

We used a 50%,20%, 30% split for training validation, testing respectively for our model. To further simulate real world conditions we added data augmentation to our model.This allows our model to perform better in difficult scenarios like low lighting wrong camera alignment etc. After experimenting and testing with a lot of different kind of augmentations we found the following augmentations gave us best results were random zoom that generates extra images that are zoomed in randomly upto 20%. In the same way we added random crop upto 20% and also random brightness for varied conditions. Augmentations effectively increases the size of our dataset and also makes sure the model works better in unknown conditions.

We trained our model on a computer with a 1070 ti.Our model was trained for 4000 steps which took 3hrs to validate accuracy of 95.6%.

IV. Results

In this section we will discuss about the results we have received after implementing the system. Our model received 95.6 percent accuracy in our validation test dataset and along with training mobilenets we also trained an inception v2 model and resenet-50 and vgg-16 model to compare the performance we received on the raspberry pi. These results are an average of all the 5 runs done on the same raspberry pi. Thus after looking at the results we can come at a conclusion that the mobilenet is the most ideal neural net model for our use case.

TABLE 2
Comparison of different models

Sr.no	Model name	CPU usage		FPS	Accuracy
1.	Inception V2	90%		1	98%
2.	Mobilenet V2	66%		6	95%
3.	Resnet-50	91%		2	97%
4.	VGG60	93%		1	96%

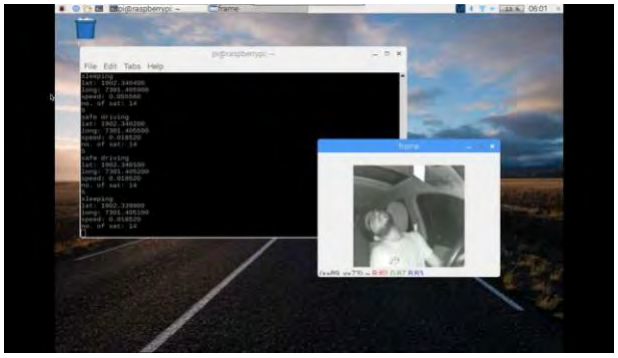


Fig. 4. Raspberry Pi screen

Although the model works as its intended to it was noticed by us that its very dependent on the driver who is driving and to remove that factor from the results we must record the dataset with a multitude of drivers.

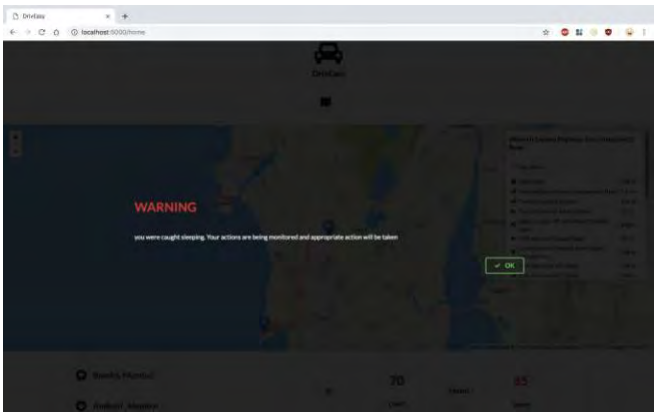


Fig.5 Web site screen for offense

V. Conclusion and Future Scope

On successful implementation the system will provide a robust and efficient method to monitor driver activities and thus prevent accidents that occur due to distracted driving, overuse of mobile phones while driving, texting on phone, drowsiness, sleeping etc. When such a system is in place it will enforce the drivers to be more careful and drive responsibly which will prevent loss of lives and will promote a safer driving experience for other drivers on the road

Once applied over a large number of vehicles the system can also be used to create a network of vehicles to share important information. In the future this network will be able to collect huge amounts of data and this data can be used to plan routes better. More over since we have deployed a hardware platform more and more features can be added in due time. Features such as facial recognition for authentication and various kinds of analysis can be done using the data of our platform. The algorithm in the proposed system relies on a neural network to detect driver's activity which performs well but an object detection approach can be used to detect specific distracting objects which will theoretically perform even better than standard neural net approaches.

References

[1] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner"Gradient based learning applied to document recognition"

[2] Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian sun "Deep Residual Learning for Image Recognition"

[3] Y.Dong, Z.hu, K.Uchimura, N.Murayama "Driver Inattention Monitoring System for Intelligent Vehicles: A Review" IEEE Transactions on Intelligent Transportation Systems, VOL. 12, NO. 2, JUNE 2011

[4] F.Vicente , Z.huang , X.Xiong , F.Torre, W.Zhang , D.Levi "Driver Gaze tracking and Eyes off the Road Detection System", IEEE Transactions on Intelligent Transportation Systems, Vol: 16 , No.4 Aug 2015

[5] A.Koesdwiady , S.Bedawi , C.Ou , F.Karray "End-To-End Deep learning for Driver Distraction Recognition", Image Analysis and Recognition, 14th International Conference, ICIAR 2017, July 2017

[6] C. Szegedy et al., "Going deeper with convolutions" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015.

[7] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H. (2017). "MobileNets: Efficient Convolutional Neural Net

[8] Ching-Hua Weng, Ying-Hsiu Lai, Shang-Hong Lai,"Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network", In Asian Conference on Computer Vision Workshop on Driver Drowsiness Detection from Video, Taipei, Taiwan, Nov. 2016.

Prediction of Next Mobile Access Gateway in Proxy Mobile IPv6 on the basis of Mobile Node Movement

Dr. Nitesh M. Tarbani1, Dr. Abrar S. Alvi2, Dr. G. R. Bamnote3 and Dr.V.S.Gulhane4
Sipna College of Engineering, Amravati, Maharashtra, India
PRMIT&R, Badnera-Amravati, Maharashtra, India
ntarbani@gmail.com,asalvi@mitra.ac.in,grbamnote@mitra.ac.in,vijaygulhane27@gmail.com

Abstract.-**The Proxy Mobile IPv6 (PMIPv6) is a network-based mobility man-agement protocol which counts on MIPv6 signaling and reuse home agent's ac-tivity through a proxy mobility agent in a localized network. In any type of wireless communication, handover delay should be as less as possible. Tremendous work has been done to minimize the handover delay in PMIPv6 and many solutions have been proposed by the many researchers. Almost all of the solu-tions have one thing in common that the authentication information of the MN should be sent to new MAG in advance and to send information to new MAG, it is mandatory to anticipate the new MAG. This paper proposes an algorithm to predict new MAG and also provides the results achieved by simulating the pro-posed algorithm on NS-2.29.**

Keywords:- Pmipv6, Mobile node, AAA server, MAG

I. INTRODUCTION

In the current period of wireless technology, mobility is gaining popularity for enabling users to access resources while roaming. There are mainly two models which provide mobility by the use of mobile IP: Network Based Mobility Management & Host Based Mobility Management. In Network based mobility management, the func-tions related to mobility management reside only in network entities and it does not need involvement of MN in the signaling or management of their movement. That means, MN remains completely unaware of its mobility. On the other hand, active participation of MN is needed in host based mobility management. Therefore, an additional software stack should be installed in MN. Thus, host based mobility management requires updating current mobile node to fit into the network whereas there is no such requirement in network based mobility mangement. For this reason, InternetEngineering Task Force (IETF) has standardized Proxy Mobile IPv6 (PMIPv6) as the network based mobility management protocol.

In PMIPv6, the mobile node is privileged to roam within that localized network provided by the PMIPv6 domain. Thus, the network administrator may have overall knowledge-set of complete network, also the resources/assets are made gettable at every PoAs and their adjoining PoAs as well. In PMIPv6, there are three network entity which take care of functions related to mobility management: Local Mobility Anchor (LMA), Mobile Access Gateway (MAG) & Authentication, Authorization, and Accounting (AAA) Server. To use the network, MN needs to get registered to LMA and then MN can access network through MAG. While moving in the network, MN may get disconnected from current MAG and get connected to the new MAG. Each time, when MN gets connected to MAG, it verifies authentication of MN fom LMA which increases hand over delay significantly.To reduce hand over delay, many researchers have proposed that current MAG should send authentication information of MN to new MAG. This is possible when new MAG should be known to current MAG. This paper focuses on the implementa-tion of algorithm that predicts new MAG with the help of MN movement. This paper also presents results ach

PMIPv6 has three main entities: LMA, AAA server & MAG. The LMA maintains the information to reach MNs and all the packets sent to or sent from MN has to pass through LMA only. In PMIPv6, the function of mobile access gateway is to do the mobility management functions and it is installed in the link to which the MN gets attached. When MAG gets MN's router solicitation message in its area, MAG verifies MN's legitimacy using MN's policy profile which is maintained at AAA server.

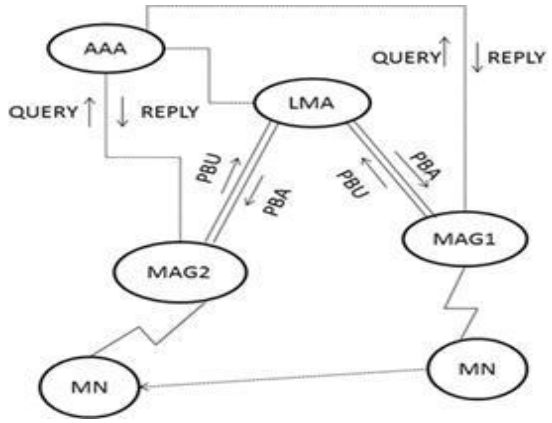


Fig. 1. Working of PMIPv

To use PMIPv6 network, MN first needs to get registered to LMA. Router solicita-tion message of MN is received by MAG. MN should include MN-ID in router solici-tation message. After receiving this message by MAG, to know LMA address, MAG uses Query Message to ask AAA server about the LMA address of MN.

When the QUERY message is received at AAA server, AAA server searches for the policy profile of the MN using MN-ID present in QUERY message. If the policy profile is not found at AAA server that means MN is not allowed to use network. If the policy profile is found at AAA server, AAA server needs to send the REPLY message carrying LMA address of MN to MAG.

When the REPLY message is received at MAG, MAG gets LMA address present in the REPLY message. After this MAG needs to create Proxy Binding Updates mes-sage and send it to the LMA to notify LMA about reachability state of MN.

When LMA receive PBU, it creates bi-directional tunnel between itself and MAG. After this, it creates message named Proxy Binding Acknowledgement which carries the MN's address prefix.

MAG sends Router Advertisement message after receiving PBA from LMA. When MN receives Router Advertisement message, it will be finally able to communicate in the network.

II. LITERATURE REVIEW

Academic and industries have done widespread research work to reduce Hand off delay in host based mobility management protocols and network based mobility man-agement protocol.

D.Johnson et al proposed MIPv6 which describes functioning of MIPv6. MIPv6 is host based mobility model in which each mobile node is always identified by its home address, regardless of its current point of attachment to the Internet [1]. Mobile IPv6 allows a MN to be connected to the Internet while travelling from one Access point to another, a process is called as handover. During handover, there is a time period when the MN cannot communicate to the internet due to various operations such as link switching, IP protocol tasks, movement detection, new Care of Address configuration, and Binding Update etc. Above mentioned handover delay as a result of typical Mo-bile IPv6 operations is mostly not tolerable to real-time traffic.

To reduce this handover delay R. Kodali et al. has proposed Fast Handover for Mobile IPv6. In this various factors are focused to reduce handover delay such as how to enable a MN to send packets as soon as it finds a new subnet link and how to deliv-er packets to a MN as soon as it becomes reachable [2].

H. Soliman et al also proposed Hierarchical Mobile IPv6 Mobility Management in which handover delay has been reduced in HPMIPv6 by introducing one more level in hierarchy and one more MIPv6 node called as Mobility Anchor Point [3].

Similarly, in network based mobility management, several works have been per-formed to minimize handover delay. V. Deverapalli et al. proposed a network based mobility management protocol namely "Proxy Mobile IPv6". Proxy Mobile IPv6 protocol is intended for providing network-based IP mobility management support to

a mobile node, which does not require the participation of the MN in IP mobility re-lated operations. Local Mobility Anchor (LMA) and the Mobile Access Gateway (MAG) are the core functional entities in the PMIPv6. In PMIPv6, AAA server can be used to store policy profile of MNs to keep track of genuine MN [5].

Ming-Chin Chuang et al. proposed FH-PMIPv6, defines predictive handoff mode in which the next MAG is informed about the handover information in advance. Both the phases authentication and registration are performed simultaneously to decrease handoff delay, in FHPMIP [9].

Ahmad Rasemy et al. have combined handoff with route optimization (O-PMIPv6) into their proposed protocol called Optimized Proxy MobileIPv6. Optimized signaling cost, network utilization and handover delay, the performance of PMIPv6 and F-PMIPv6 is improved by the proposed protocol [10].

Geumsan Jo et al. have proposed method in which an MN movement history is maintained at Local Mobility Anchor (LMA) and proposed method compares the current location of the MN with the prior location and using which anticipate the nMAG [12].

III. PROPOSED METHODOLOGY

As discussed earlier, to reduce handover delay in PMIPv6, many researchers have proposed that current MAG must send authentication information of MN to new MAG. Now, each MAG is surrounded by many other MAGs. It means that, current MAG should be able to predict the new MAG. Following is the algorithm proposed here to predict the new MAG.

In this work, the area around current MAG is divided into eight different parts. Once the MN is about to go out of coverage area of current MAG, the MAG will find out the current location of MN using GPS. Upon knowing current location of MN, MAG can find out the nearest MAG to the MN using following Haversian formula.

$a = \sin^2(\Delta\varphi/2) + \cos\varphi1 \cdot \cos\varphi2 \cdot \sin^2(\Delta\lambda/2)$

$c = 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{(1-a)})$

$d = R \cdot c$

where φ is latitude, λ is longitude, R is earth's radius (mean radius = 6,371km);

Pseudo JavaScript Code :

```
var R = 6371000; // metres

var φ1 = latitude1.toRadians();

var φ2 = latitude2.toRadians();

var Δφ = (latitude2-latitude1).toRadians();

var Δλ = (longitude2-longitude1).toRadians();

var a = Math.sin(Δφ/2) * Math.sin(Δφ/2) +

Math.cos(φ1) * Math.cos(φ2) *

Math.sin(Δλ/2) * Math.sin(Δλ/2);

var c = 2 * Math.atan2(Math.sqrt(a), Math.sqrt(1-a));

var d = R * c;
```

As the proposed methodology has been simulated in the NS-2.29, following formu-la is used to find out distance between MAG and current location of MN.

$= \sqrt{(\ 1 -\ 2)2 + (\ 1 -\ 2)2}$

Where, (X1,Y1) and (X2,Y2) are locations of MN and MAGs respectively.

Following is the pseudo code for finding the nearest MAG to MN current location.

```
if(dist >threshold && mniflag==0)

{

for(i=0;i<n;i++)

{

dx=xpos-amxpos[i];

dy=ypos-amypos[i];

dxs=dx*dx;

dys=dy*dy;

dist=dxs+dys;

if(dist<min)

{

min=dist;

mini=i+2;

}

}

mniflag=1;

}
```

Where, dist is the distance of MN from current MN,

Threshold is used to decide whether MN is likely to leave coverage area or not, xpos is x coordinate of MN, amxpos is the array which contains the MAG addresse

5 RESULT & ANALYSIS

For simulation, one AAA server, one LMA, nine MAGs, one corresponding node (CN) & one MN are considered and CN needs to send data to MN.

Topology is simu-lated in such a way that MN will get connected to MAG which is surrounded by 8 other MAGs. When MN is about to get disconnected from current MAG, current MAG uses above proposed methodology to predict the next MAG. Table 1 shows the configuration of other necessary parameters for the simulation:

Simulation tool	NS-2.29
Simulation time	20sec
Number of LMAs	1
Number of MAGs	9

Table 1. Simulation Parameters

Packet size	1000 bytes
Routing protocol	PMIPv6 with AAA
Traffic type	CBR
Destination of MN	Each time, different destination is set

The correctness of proposed methodology is evaluated on the basis of predicted new MAG and actual new MAG. The TCL script is run again and again with new destination of MN so that each time, the new MAG for MN should be different. Each time, readings of predicted new MAG and actual new MAG are noted down. The readings, in which the predicted new MAG and actual new MAG are same, show that the prediction was correct whereas the readings, in which the predicted new MAG is not same as actual new MAG, show that the prediction was not correct. To evaluate the correctness of proposed methodology following formula was used.

Where, CPr is number of correct predication got and TPr is total number of predictions taken.

In all, total 500 times code was run with different destinations. Out of which 437 times the predications were correct and 63 times predictions were incorrect. So, ac-cording to the formula mentioned above, the percentage of correctness of proposed methodology is 84.7%.

Handoff delay also was observed for correct predictions and incorrect predictions. Handoff delay is nothing but the time difference between the time at which the MN got disconnected from current MAG and connected to new MAG. Following figure explains the handoff delay.

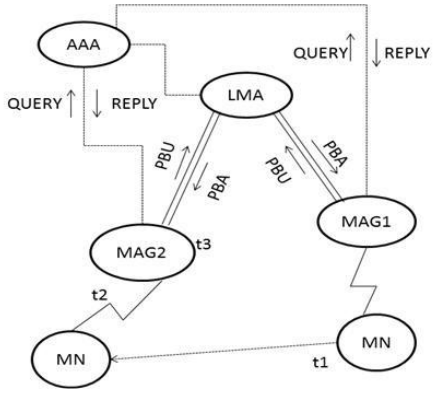


fig. 2. Handoff delay

In the above fig, t1 is the time at which MN got disconnected from current MAG (MAG1), at time t2, MN discovered the new MAG (MAG2). Though MN discovered

Situation	Handoff Delay
Without Prediction	8.044ms
Correct Prediction	**4.024ms**
Incorrect Predictions	8.044ms

new MAG, MN cannot use the network until new MAG receives PBA from LMA. At time t3, MAG2 receives PBA from LMA and at t3, MN now can use the network. So, handoff delay, in this case can be given as follows.

Handoff delay = t3-t1;

Following table shows the handoff delay observed in various situations.

Table 2. Handover Delay

The table shows that three scenarios were observed. First one, when prediction algorithm is not used, handoff delay found in this scenario is 8.044ms. Second scenario, when prediction algorithm is

used and predicted MAG was the same as of new MAG, in this scenario the handoff delay observed was 4.024ms (almost half of the previous case). In third scenario, the predicted MAG was not same as new MAG to which MN got connected. In this case also handoff delay observed was 8.044ms. Moreover, to find correctness of implemented algorithm, the code is run against 200 distinct desti-nation addresses and out of 200 runs, 184 times, the predicted address, found to be correct. This allowed us to conclude that the correctness of algorithm is 92%.

6.CONCLUSION

The PMIPv6 is one of the promising network based mobility management algo-rithm. However, the handover delay is the place where this algorithm can be opti-mized more and made more prominent. Many research works have been done in this field, to reduce handoff delay of PMIPv6. The back bone of reducing handoff delay is to send authentication information of MN to new MAG. This requires prediction of new MAG to which MN is going to connect. This paper proposed an algorithm to predict the new MAG and also presented the result achieved. It is observed that use of prediction algorithm leads to reduction in handoff delay to almost half. Moreover, to find correctness of prediction algorithm, the algorithm is run 200 times with distinct destination addresses of MN. The correctness of algorithm is found to be 92%.

Reference

1. D. Johnson, C. Perkins, "Mobility Support in IPv6", RFC3775, June 2004.
2. R. Koodli, “Fast Handovers for Mobile IPv6,” Internet Engineering Task Force Request for Comments - 4068, July 2005.
3. H. Soliman, C. Castelluccia, K. El Malki, and L. Bellier, “Hierarchical Mobile IPv6 Mo-bility Management (HMIPv6),” Internet Engineering Task Force Request for Comments - 4140, August 2005.
4. Linoh A. Magagula, H. Anthony Chan “802.21 Optimized Handover Delay for Proxy Mobile IPv6”, Military Communications Conference IEEE, May 2008.
5. V. Devarapalli, Wichorus, K. Chowdhury , Proxy Mobile IPv6, RFC 5213, August 2008
6. Farhana F. Liza, Dr. Wenbing Yao “Implementation Architecture of Proxy Mobile IPv6 Protocol for NS2 Simulator software”, 09th , International Conference on Communication Software and Networks IEEE, July 2009.
7. Meng-Hsuan Lin, Whai-En Chen*, Chao-Hsi Huang, “HF-PMIPv6: An Enhanced Fast Handovers for Network-based Mobility Management”, International Conference on Ad-vanced Infocom Technology IEEE, July 2011.
8. N.M.Tarbani, Chandavarkar B.R. “Implementation of AAA Server for PMIPv6 in NS-2” presented in International Conference on Parallel, Distributed Computing Technologies and Applications, PDCTA 2011, Tirunelveli, India, September 23-25, 2011 Pg.no 523-531 (Online ISBN 978-3-642-24037-9 Springer).
9. Ming-Chin Chuang and Jeng-Farn Lee, “FH-PMIPv6: A Fast Handoff Scheme in Proxy Mobile IPv6 Networks”, International Conference on Consumer Electronics, Communica-tions and Networks IEEE, November 2011.
10. Ahmad Rasemy, Christian Makaya, and Marc St-Hilairey, “O-PMIPv6: Efficient Hando-ver with Route Optimization in Proxy Mobile IPv6 Domain”, International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) IEEE, De-cember 2012.
11. Zhiwei Yan and Jong-Hyouk Lee, Senior Member, “State-Aware Pointer Forwarding Scheme With Fast Handover Support in a PMIPv6 Domain”, IEEE Systems Journal, VOL

7, No. 1, March 2013.

12. Geumsan Jo,Hyun Jung Choe and Hyunseung Choo,“Predictive Handover Scheme Using Mobility History in PMIPv6” , Research in Adaptive and Convergent Systems 13,ACM October 2013.
13. N.M.Tarbani, Dr. A. S. Alvi "Use of Global Positioning System to Track Movement of Mobile Node in Proxy Mobile Internet Protocol Version 6" presented in International Con-ference on Computer and Communication Technologies" Pg. No. 103-108 September 2015 (Online ISBN 978-81-322-2517-1 Springer)
14. N.M.Tarbani, Dr. A. S. Alvi, Dr. G. R. Bamnote, Dr. S. V. Pattalwar "Prediction of next Mobile Access Gateway by Tracking Locations of Mobile Node" Presented in Internation-al Conference on Information and Communication Technology for Competitive Strategies March 2016 (ISBN: 978-1-4503-3962-9 ACM)

Surveillance System for Intruder Detection using Facial Recognition

Shaikh Mohammed Umraan,Dr. Deepali Vora,Dr. Abhishek Anurag
Department of Information Technology,Ph.D. (CSE) Pune, India
Vidyalankar Institute of Technology
Mumbai, India
umraans@gmail.com,deepali.vora@vit.edu.in,anuragabhishek@gmail.com

Abstract — **Facial recognition system is used widely to identify and verify the person's face from image or video source. With the continuous expansion of the surveillance system, surveillance cameras not only bring convenience, but also produce a massive amount of monitoring data, which poses huge challenges to storage, analytics and retrieval. The smart monitoring system equipped with intelligent video analytics technology can monitor as well as pre-alarm abnormal events or behaviors. Here, propose system will detect the intruder and inform the security within seconds. The NVIDIA Jetson Nano board will be used to compute Convolutional Neural Network algorithm for the facial recognition process. The basic idea will be to use this system where a database can be stored of the existing faces. The system will then take the data from the surveillance camera and run facial recognition algorithm on it. It will match all the faces with the ones already stored in the database and if it finds any face which is new, it will send an alert to the security personnel. This will help to increase the security of the place where there are many people gathered at a time for example schools, colleges, universities, etc.**

Keywords—Facial Recognition, Biometric, NVIDIA Jetson Nano, Surveillance

I. INTRODUCTION

Traditional video surveillance can only provide simple functions such as video capture and storage. It cannot automatically alarm for any intruders. In order to find any intrusion in real time, monitor personnel need to constantly observe the video. In this case, the monitor faces dozens of surveillance video images, which is easy to fatigue. It may not be able to respond in time to intrusion due to lack of concentration, and loose key information in the video. [1]
In addition, because a large amount of surveillance data, video needs to be stored for months or years, it will result in a large storage cost. Therefore, the intelligent video surveillance system is urgently needed to assist the monitoring personnel to use the intelligent detection technology to process, analyze and understand the video signal while retaining the original video key information, andautomatically detect the target and location information without manual intervention. [1]
In the event of an intruder, an alarm is issued in time to effectively assist the monitoring personnel. The traditional moving target detection method can only detect frames with moving targets, but cannot understand the semantic content in the video. With the development of deep learning, advanced target detection, semantic understanding, and instance segmentation techniques are emerging, which realizes the semantic understanding of video content and improves the accuracy. In recent years, with the advancement of video analysis technology, intelligent video surveillance systems have developed rapidly, and a large number of representative technical achievements and applications have emerged. [2] With the help of embedded video analysis algorithms, the intelligent surveillance camera can identify the abnormal behavior of the monitoring site without manual intervention. Typical anomalous behaviors include entering and leaving an area, moving quickly, gathering, and so on.
The Intelligent Video Analytics Model (IVAM) also known as Human Object Detection Model will be used. The block diagram below shows how the model process flows.

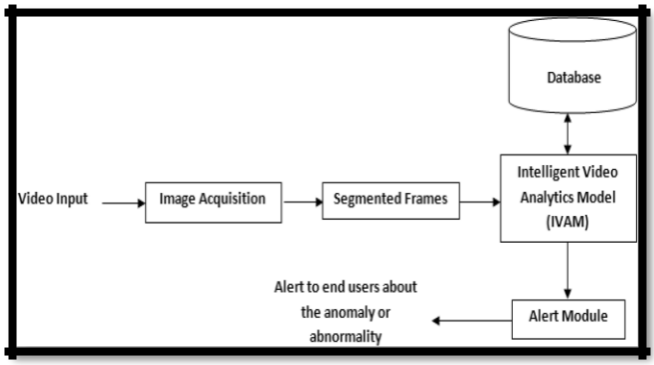


Fig 1 (IVAM Block Diagram)

Fig. 1.1 shows the model for proposed surveillance system. The basic flows runs as, first the data is fed as video input. From the video input the image will be acquired and each frame will be segmented. The IVAM will be applied on the segmented frames. In this step the feature extraction will be done. The data from the database will be matched to the output from IVAM. If it doesn't match then the alert will be triggered to notify the security personnel. Once the intruder is detected the image of the intruder will be snapshot along with the timeframe and will be saved in the database as evidence. [3]

II. LITERATURE SURVEY

Video monitoring and analysis for facial recognition problem has been studied by many researchers and different techniques/ approaches have been suggested in the Table 1.

III. PROPOSED SYSTEM

To address the above issues in facial recognition, we propose a better surveillance system using NVIDIA Jetson Nano instead of Raspberry Pi. Jetson Nano have much higher AI performance than Raspberry Pi. It has higher RAM and GPU which provides faster results compared to the Raspberry Pi. Jetson Nano board runs multiple neural networks simultaneously hence it provides accurate results

in much less time. The proposed system will be a much better surveillance system which will run on the Jetson Nano. The proposed system will be evaluated on the basis of efficiency, accuracy and runtime. The OpenCV library will be used on the Jetson Nano to run the system. Fig. 2 shows the process flow for the facial recognition system. The video input will be fed to system.

Face Detection: First the system will detect all the faces from the video feed. Here the Multi Task CNN algorithm will be used. This algorithm runs 3 neural networks. The first network will detect the face by scaling the image to a particular size. Jetson Nano supports DeepStream SDK and samples. These will be help in face detection.

Face Alignment: The second neural network will align the face and add a padding to the faces which are partially outside the screen/ bounding box. The third network will provide the final result i.e. the aligned face.

Feature Extraction: FastICA algorithm to be used. It will add the geometric feature to the face for independent component analysis.

Feature Matching: Again the Multi Task CNN algorithm will be used to match the features with the existing ones in the database.

Results: Finally, if the features doesn't match then the alarm is triggered to the security.

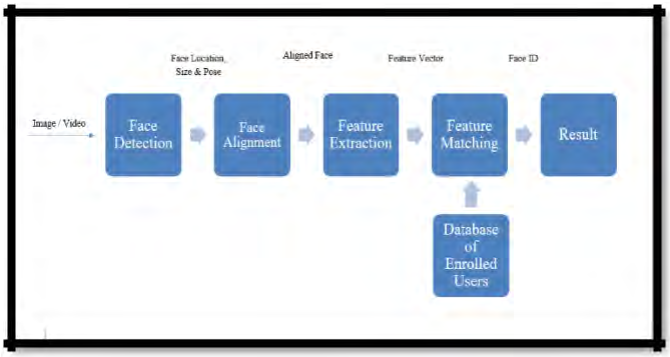


Fig 2 (Block Diagram of Proposed System)

The proposed system uses Jetson Nano board by Nvidia which will provide better performance as compared to the traditional Raspberry Pi boards. Following table shows the comparative study between the Jetson Nano board and Raspberry Pi board based on 13 parameters.

Table 2 shows that Jetson Nano has higher AI performance which is useful will applying deep learning algorithms and which will give a better edge at running the proposed surveillance system. The high powered CPU, GPU and RAM also helps in faster computations of complex algorithms as well as it helps to run multiple neural networks simultaneously. The video output of the Jetson Nano has 4K resolution thus providing clear image. The only constraint which goes against the proposed hardware is the price of the
hardware but given the specifications and computation power it is worth the money.

Table 2 (Comparative Study)

	Jetson Nano Dev Board	Raspberry Pi 3A+	Raspberry Pi 3B+
AI Performance	472 GFLOPS	21.5 GFLOPs (est*)	21.4 GFLOPs (est*)
CPU	1.4 GHz 64-bit Quad-Core ARM Cortex-A57 MPCore	1.4 GHz 64-bit Quad-core ARM Cortex-A53	1.4 GHz 64-bit quad-core ARM Cortex-A53
GPU	128-Core Nvidia Maxwell	Broadcom VideoCore IV	Broadcom VideoCore IV
RAM	4GB LPDDR4	512MB LPDDR2 SDRAM	1GB LPDDR2 SDRAM
GPIO Header	40-pin	40-pin	40-pin
Board Dimensions	100 X 79mm	65 X 56mm	85 x 56mm
Wireless	None	Dual-band 802.11ac wireless LAN, Bluetooth 4.2/BLE	Dual-band 802.11ac wireless LAN, Bluetooth 4.2
Ports	4x USB 3.0, wired ethernet 10/100/1000Mbps	1 USB 2.0	4 USB 2.0, Wired Ethernet up to 330 Mbps
Multimedia	2160p30 (H.264)	1080p30 (H.264)	1080p30 (H.264)
Video Output	HDMI, Display Port (4K)	HDMI, Display Serial Interface (DSI)	HDMI, Display Serial Interface (DSI)
Camera Serial Interface	YES	YES	YES
M.2 Key E Slot	YES	NO	NO
Price	$99	~$25	~$35

IV. CONCLUSION

The above proposed surveillance system for the facial recognition promises a much efficient and highly accurate in giving the results. The use of the Jetson Nano will be critical in the performance of the entire system. Due to the use of Jetson Nano, the entire process will run much faster and multiple neural networks will run simultaneously. Moreover the use of Multi Task CNN algorithm for face detection, alignment and feature matching will also provide much better results One of the major advantages of the proposed system is that Jetson Nano uses the same CUDA and OpenCV libraries along with its own Jetpack SDK's. Along with these, Nvidia's TensorRT will also help for high performance deep learning inference. So using the existing libraries it will be much easier to design an efficient system which will work much faster and more accurately.

References

[1] Kyu Beom Lee, Hyu Soung Shin," An application of a deep learning algorithm for automatic detection of unexpected accidents under bad CCTV monitoring conditions in tunnels", IEEE, 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, DOI: 10.1109/Deep-ML.2019.00010

[2] Ishani Mondal, Sombuddha Chatterjee, "Secure and Hassle-Free EVM Through Deep Learning Based Face Recognition", IEEE, 2019 International Conference on ML, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India DOI: 10.1109/COMITCon.2019.8862263

[3] Liyun Zhuang, Yepeng Guan, "Deep Learning for Face Recognition under Complex Illumination Conditions Based on Log-Gabor and LBP", 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China DOI: 10.1109/ITNEC.2019.8729021

[4] Yun Xia Liu, "Intelligent Monitoring of Indoor Surveillance Video Based on Deep Learning", IEEE, 2019 21st International Conference on Advanced Communication Technology (ICACT), Korea DOI: 10.23919/ICACT.2019.8701964

[5] M.Ali et al, "Edge Enhanced Deep Learning System for Large Scale Video Stream Analytics", 2018 IEEE 2nd International Conference on Fog and Edge Computing (ICFEC), Washington DC, USA DOI: 10.1109/CFEC.2018.8358733

[6] Yi Shan, "ADAS and Video Surveillance Analytics System using Deep
Learning Algorithms on FPGA", IEEE, 2018 28th International Conference on Field Programmable Logic and Applications (FPL), Dublin, Ireland DOI: 10.1109/FPL.2018.00092

[7] A. Balasundaram, C. Chellappan, "An intelligent video analytics model for abnormal event detection in online surveillance video", Springer,
Journal of Real Time Image Processing, pp 1-16, 2018 DOI: https://doi.org/10.1007/s11554 -018-0840-6

[8] Abdulkadir Sengur et. Al., "Deep Feature Extraction for Face Liveness Detection", IEEE, 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Turkey DOI: 10.1109/IDAP.2018.8620804

[9] Camille Bailas, Mark Marsden, Dian Zhang, "Performance of Video Processing at the Edge for Crowd Monitoring Applications", IEEE,

2018 IEEE 4th world Forum on Internet of things(WF-IoT),Singapore DOI : 10.1109/WF-IoT.2018.8355170

[10] Xiujie Qu, Tianbo Wei, Cheng Peng, Peng Du, "A Fast Recognition System Based on Deep Learning", IEEE, 2018 11th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, China DOI: 10.1109/ISCID.2018.00072

[11] Milena Napiorkowska, David Petit, Paula Marti, "Three Applications of Deep Learning Algorithms for Object Detection in Satellite Imagery", IEEE, IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, DOI: 10.1109/IGARSS.2018.8518102

[12] Muhammed Usman Yaseen, Ashiq Anjum, Omer Rana, Nicholas

Antonopoulos, "Deep Learning Hyper Parameter Optimization for Video Analytics in Clouds", IEEE, IEEE Transactions on Systems, Man, and Cybernetics: Systems,Pp : 253 – 264,DOI: 10.1109/TSMC.2018.2840341

[13] Xukan Ran, Haoling Chen, Xiaodan Zhu, Zhenming Liu, Jiasi Chen,

"DeepDecision: A Mobile Deep Learning Framework for Edge Video Analytics", IEEE, IEEE INFOCOM 2018 - IEEE Conference on

Computer Communications, Honolulu, HI, USA, DOI:10.1109/INFOCOM.2018.8485905

[14] Ali Elmahmudi, Hassan Ugail, "Experiments on Deep Face Recognition using Partial Faces", IEEE, 2018 International Conference on Cyberworlds (CW), Singapore, Singapore,DOI: 10.1109/CW.2018.00071

[15] Onur Can Kurban, Ahmet Bilgic, "A Multi-Biometric Recognition System Based on Deep Features of Face and Gesture Energy Image",

IEEE, 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, Poland, DOI: 10.1109/INISTA.2017.8001186

[16] Khalid Tahboub, David Guera, Amy Reibman, Edward Delp, " Quality Adaptive Deep Learning for Pedestrian Detection", IEEE, 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China DOI: 10.1109/ICIP.2017.8297071

[17] Poonam Sharma, R. N. Yadav, K. V. Arya, "Face Recognition from Video using Generalized Mean Deep Learning Neural Network",

IEEE, 2016 4th International Symposium on Computational and Business Intelligence (ISCBI), Olten, Switzerland, DOI: 10.1109/ISCBI.2016.7743283

[18] Atika Burney, Tahir Q, Syed, "Crowd Video Classification using CNN", IEEE, 2016 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, DOI: 10.1109/FIT.2016.052

[19] https://motchallenge.net/data/MOT17/ test data

[20] https://developer.nvidia.com/embedded/jetson-nano-developer-kit,

"Jetpack SDK", accessed on 16th December 2019 at 7.45pm

[21] https://towardsdatascience.com/how-does-a-face-detection-program-work-using-neural-networks-17896df8e6ff, "Face detection program using neural networks" accessed on 17th December 2019 at 4.30 pm

Table 1 (Literature Survey)

Authors	Abstract	Algorithms	Inference from Existing System
Kyu Beom Lee, Hyu Soung Shin [1]	Proposes a new process of ODTS by combining deep learning-based object detection network and object tracking algorithm and it shows dynamic information of an object for a specific object class can be obtained and utilized.	R-CNN, Object Tracking Algorithm	This system adds CADA that discriminates every cycle based on dynamic information of the car objects. As a result of experimenting with the image containing each accident, it was possible to detect the accidents within 10 seconds.
Ishani Mondal, Sombuddha Chatterjee [2]	This paper proposes face recognition involves a large number of challenges when it comes to deal with the visual analysis. Thus, this development or approach plays a significant role in security applications such as legal documents identification, identification of terrorists in public places.	CNN Algorithm	The proposed method has been used in real-time, runs perfectly on the standard benchmark dataset as well as the custom dataset prepared by us. The Performance has been better compared to the traditional approaches just because of the deeper architecture of the convolutional neural networks.
Liyun Zhuang, Yepeng Guan [3]	A novel method based on deep learning to solve the adverse impact imposed by illumination variation in the face recognition is proposed in this paper.	Log-Gabor Algorithm, LBP Algorithm	Experimental results have shown that the proposed method has superior performance in face recognition by comparing with some state-of-the-arts.
Yun Xia Liu, Yang Yang, Aijun Shi, Peng Jigang [4]	This paper proposes a method based on Mask R-CNN for intelligent monitoring of indoor surveillance video.	R-CNN Algorithm	It uses advanced target detection and instance segmentation technology in deep learning to realize the semantic understanding of video content, which not only effectively retains
Authors	**Abstract**	**Algorithms**	**Inference from Existing System**
			the key information in the original video.
M. Ali et. al. [5]	In this paper, an edge-based system for deep learning is proposed for efficient and large-scale video stream analytics. Using the infrastructure, an object recognition scenario was implemented.	CNN Algorithm	Results on an object recognition scenario show 71% efficiency gain in the throughput of the system by employing a combination of edge, in-transit and cloud resources when compared to a cloudy-only approach.

Yi Shan [6]	This paper proposes Deep learning algorithms, such as CNN could provide high accuracy for great number of applications including ADAS and video surveillance analytics. Considering processing speed and energy efficiency, FPGA is a good hardware to construct customized CNN solution.	CNN Algorithm	This system processes 16 channels of continuous input video with the resolution of 1080p. Two functionalities could be easily switched by just clicking a button in this live demo: one ADAS system for vehicle, non-motorized vehicle, and pedestrian detection, tracking, and attributes analytics; and the other video surveillance system for face detection and recognition. The FPGA used is Xilinx MPSoC ZU9, and the whole board including this FPGA only cost about 50 Watts with Peak performance at 5.6 TOPS.
A. Balasundaram, C. Chellappan [7]	The main objective of this paper is to design and implement a novel intelligent video analytical model as a human object detection method for surveillance video.	Intelligent Video Analytics Algorithm	It is found and concluded that the IVA model outperforms most of the existing approaches in terms of object detection and classification with less error. From the results, it is observed that the IVA model has achieved 99.77% and 98.19% accuracy in correctly classifying the normal and abnormal frames from benchmark dataset and custom dataset, respectively. Similarly, 0.23% and 2.03% of normal and abnormal frames are classified incorrectly from benchmark dataset and custom dataset, respectively.
Abdulkadir Sengur, Zahid Akhtar, Yaman Akbulut [8]	Face recognition systems are now being used at industrial level throughout the world. Despite the great advances, counteracting face spoofing attacks has yet proven to be a challenging task. Therefore, in this paper, as a first step we have investigated two well-adopted CNN models for face PAD without any fine-tuning.	CNN Algorithm	Investigation of two well-adopted CNN models for face PAD without any fine-tuning, which is contrary to the prior studies. Experimental results on two publicly available databases are presented. The future work includes a comprehensive analysis of several state-of-the-art deep learning models for face PAD.
Camille Bailas, Mark Marsden,	In the early age, smart city applications were relying on a cloud centric approach. Thus, in this paper it is explored that video analytics at the edge using a Dell	Crowd Density, ResNet50,	The algorithm can process more than one frame per second where most CCTVs only capture one image per second which make them suitable for near real-time crowd

Dian Zhang [9]	Edge Gateway 5000. They can process more than one frame per second where most CCTVs only capture one image per second.	Crowd Counting	monitoring. However, with the algorithm complexity increasing, the size of the frame highly impacts the performance. It can increase by a factor of 5 the processing time at the
Authors	**Abstract**	**Algorithms**	**Inference from Existing System**
			edge. Carefully choosing the input size is important as a good trade-off between the model accuracy and the runtime performance can be found saving up to 30% time.
Xiujie Qu, Tianbo Wei, Cheng Peng [10]	With the advent of the era of big data, deep learning theory has been rapidly developed and applied, especially in the field of image recognition. Based on the principle of CNN, a method of realizing face recognition on FPGA is presented in this paper.	CNN Algorithm	The speed of face recognition system is 400FPS, the recognition rate is 99.25%, and what's more, it has good robustness, which means it can complete the recognition function under most light conditions. Compared with the existing results, the indicators of the system have been significantly improved.
Milena Napiorkowska, David Petit, Paula Marti [11]	Detection of objects in images has been long used in computer vision applications in fields such as surveillance or robotics. This paper shows how one of the networks developed for the ImageNet challenge can be applied to satellite imagery for object detection using three examples: roads, palm trees and cars.	VGG Algorithm	Results show that our approach is good at finding objects that might have different colours and slightly varying shapes, which cannot be achieved as easily using more common techniques in remote sensing such as Random Forest or Support Vector Machine.
Muhammed Usman Yaseen, Ashiq Anjum, Omer Rana [12]	A system to perform video analytics is proposed using a dynamically tuned convolutional network. Videos are fetched from cloud storage, preprocessed, and a model for supporting classification is developed on these video streams using cloud-based infrastructure. A key focus in this paper is on tuning hyper-parameters associated with the deep learning algorithm used to construct	Video Analytics Algorithm	Numerous experiments on the testing dataset proved that the system is accurate with an accuracy of 0.97 as well as precise with a precision of 0.96, respectively. The system is also capable of coping with varying number of nodes and large volumes of data. The time required to analyse the video data depicted an increasing trend with the increasing amount of video data to be analyzed in the cloud. The

	the model.		analysis time is directly reliant on the amount of data being analyzed.
Xukan Ran, Haoling Chen, Xiaodan Zhu [13]	In this work, they developed a measurement-driven frame-work, Deep Decision, that chooses where and which deep learning model to run based on application requirements such as accuracy, frame rate, energy, and network data usage.	CNN Algorithm	Results suggest that Deep-Decision can make smart decisions under variable network conditions, in contrast to previous approaches which neglect to tune the video bitrates and resolution and do not consider the impact of latency on accuracy. It doesn't include object tracking to reduce the frequency of running deep Learning, generalizing the algorithm for a larger set of edge devices, and customizing the algorithm for different categories of input videos.
Ali Elmahmudi, Hassan Ugail [14]	In this paper, the use of deep learning approach for face recognition using partial face data is studied. Based on the popular CNN and using the VGGF model for extracting features from face, two different classifiers namely the cosine similarity and the linear support	VGG Face Model, CNN Algorithm	Experimental results validate that the cheek, nose, forehead and mouth have low recognition rates. On the other hand for top half of the face, right or left half and for ¾ of the face the recognition rates reach 100%. In addition, the Cosine similarity measure greatly improves
Authors	**Abstract**	**Algorithms**	**Inference from Existing System**
	vector machine for classification were utilized.		the performance of the classification when compared to the SVM.
Onur Can Kurban, Ahmet Bilgic [15]	In face recognition systems, variables such as direction of light, facial expression and reflection makes identification difficult. With biometric fusion, both safe and high performance results can be achieved	VGG Face Model	The results show that the S gesture can be used for authentication with this method. These two biometrics are fused as a score level by sum rule to increase both safety and robustness.
Khalid Tahboub, David Guera, Amy Reibman [16]	In this paper, a two-stage quality-adaptive convolutional neural network to address the problem of a changing video data-rate is proposed.	CNN Algorithm	Experimental results demonstrated that when adaptive data-rate streaming is used our proposed quality adaptive approach reduces the miss rate compared to the

			baseline detector.
Poonam Sharma, R. N. Yadav, K.V. Arya [17]	Proposed algorithm is a face recognition algorithm from video using generalized mean Deep Learning Neural Network. The performance of the proposed algorithm is tested on two most commonly used databases, i.e. PaSC and YouTube databases.	CNN Algorithm	The results proved that the proposed algorithm is better in terms of identification accuracy.
Atika Burney, Tahir Q, Syed [18]	In this paper, it is shown that the mid-level descriptors of the groups can be used to classify crowd videos and also similar accuracy can be obtained if the whole crowd is considered as a single entity.	CNN Algorithm	Results shows the additional tasks of identifying groups, computing their features and then combining them to define the crowd can be reduced to single step of computing the features of the crowd.

Secure Data Aggregation and data Transmission using HMAC Protocol in Cluster base Wireless Sensor Network

Sunny Sall and Dr. Rajesh Bansode
Department of Computer Engineering St. John College of Engineering and Management,
Department of Information Technology, Thakur College of Engineering and Technology,
Maharashtra, India
sunny_sall@yahoo.co.in ,rajesh.bansode@thakureducation.org

Abstract— **Wireless Sensor network (WSN) facing many issues in due implementation in vulnerable environments. Various researchers has already defined a systems for data transmission with WSN, still such system having high data loss issues, maximum packet delay as well as packet overhead during the data transmission. Different parameters has to consider to improve such losses like cluster network generation, data aggregation, secure data encryption and data transmission using distribution approach produces effective outcome and eliminate such issues. Moreover implementation with cluster network and selection a cluster head (CH) based on trust which produces much effective results and provides flexibility to system. This is research we proposed secure data transmission in cluster network and investigate the Quality of Service (QoS) parameters of entire execution. Initially we create different clusters with collaboration of multi sensor nodes while each node consist individual battery power as energy. According to highest energy we calculate the trust and such node defined as CH, Data Aggregation (DA) and Broadcast Tree Construction (BTC) two different techniques have been used for eliminate network lifetime or cut generation in system. In partial experiment analysis system shows improve QoS parameters like Throughput, delay, packet overhead etc respectively. It also enhance the network life because of proposed energy conservation approach.**

Keywords - Cluster Network, WSN, Data aggregation, Broadcast Tree Construction, QoS.

I. INTRODUCTION

In this era, the dagger development of mobile computing expedients that primarily embody laptops, personal digital assistants (PDAs), as well as hand-held digital devices, is driven by a revolutionary change in the computing world. Security approaches to eliminate different network attacks like wormhole attack [1] in untrusted environment. To identify intruder or threats from large traffic defined in [2]. Computing will not simply place trust within the power provided by private computers, and the idea of gift computing often arises and becomes one within the applied science society at every research hotspots [3]. Throughout this environments a path behalf of the two hosts might contains steps through one or extra nodes inside the painter. an important draw back in associate degree passing mobile ad-hoc network is finding and maintaining routes since host quality can cause topology changes [4]. Several routing algorithms for MANETs are proposed inside the writings which they differ inside the painter. New routes square measure set up and existing ones square measure modified. Basically MANET networks square portion extra susceptible to suffer from the hateful performances. This paper we proposed an wormhole attack detection and prevention approach using secure mechanism of detection such malicious behaviour. This paper we proposed an various attack detection and prevention approach using secure mechanism of detection in malicious behaviour [5] and investigate the proposed experimental analysis.

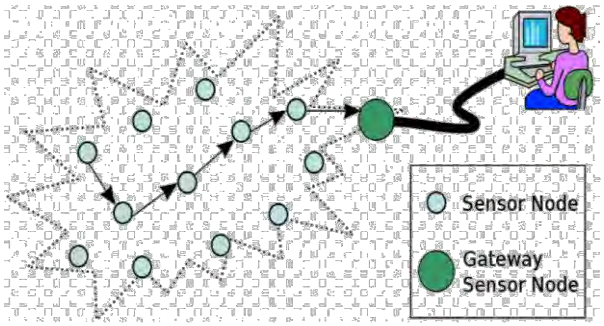


Figure 1 : Representation of a wireless sensor network [6]

The above figure 1 shows basic WSN view in sensor network, which is basically similar to wireless ad hoc networks in the intelligence that they rely on wireless connectivity and unstructured establishment of networks so that sensor data can be ecstatic wirelessly.

II. RELATED WORK

Yashpal sinh Gohil et. al. [1] proposed a true Link verifies whether or not there's an immediate Use truelink to link a node to its adjacent neighbor. The main disadvantage is that TrueLink only works on IEEE 802.11 devices that are backward compatible with an update to the computer code. An approach to trip time (RTT) is emerging to beat the problems of further hardware exploitation. The RTT is that the time taken to deliver RREQ to a supply node and to collect RREP from endpoint. A node should calculate the RTT between its neighboring nodes and itself. The malicious nodes are priced higher for RTT than different nodes. This method will determine the supply of its real and misbehavior neighbors. This detection technique is efficient only in the case of hidden attacks.

Vikaskumar Upadhyay and Rajesh Shukla [2] describes Numerous methods are planned to be used to identify wormhole attacks using a leash method for the packets. Packet leash is the technique that guards against wormhole attack. The leashes are either spatially or temporally categorized. All nodes inside the network should have data from their own position and stable synchronized clock in spatial leashes. This requires its own current location when a sender sends the data packet and UTC accepts wormhole nodes. Directional data are exchanged between source and destination during this technique. Destination by scrutiny will find the wormhole the Received signal from malicious nodes and gave directional info. The relation to the wormhole is observed if the supply signals and intermediate nodes are distinct.

Amit Kumar and Sayar Singh Shekhawat [3] describes, in Authors have criteria of contact that are primarily focused on a paradigm of research to ensure safe communication. Another such assault is a wormhole attack, in which two or many nodes are often flustered together to gain knowledge calculation and communication. A wormhole compromised network is illustrated during this post, and the research model can be used for secure communication within the network being targeted. The network model is designed to identify the secure communication node, so that connectivity is maximized. Additionally, in mostly mobile network based attack, the writers address a way of generating the safe path. The model eventually provided the optimized parameter for reconciliation communication. Results show the improved performance of model in terms of the communication throughput and reduced the loss.

Ashish Kumar Jain and Ravindra Verma [4] authors To search for aggressive or legitimate nodes inside the network, build calculative confidence values for the nodes. Authors propose a way to defend against victimizing wormhole attack combination of parameters such as energy, connection range and node buffer length. A node's confidence value is calculated backed by these parameters. This node confidence value is then compared with the threshold value of network faith. This comparison was supported and one can find out if The selected node is either malicious or valid, or is not. The planned methodology consists of two phases: first, the analysis of the network parameter and the calculation of the threshold, and second, the implementation of the protection on the routing protocol that prevails. Results area unit assessed by expanding the AODV protocol as well as evaluating the efficiency of the proposed routing protocol and matching it to the insecure AODV. The performance of the proposed confidence-based approach, based primarily on defense, is cost-effective and reasonable against comparative findings wormhole attacks in MANET.

Rajan Patel Anal Patel, Nimisha Patel [5] this Survey is The different techniques used to identify authors of wormhole attacks suggest an approach for the identification and prevention of wormhole attacks. A projected defensive approach against wormhole attack based primarily on the Hash-based Compression Function (HCF) that is really mistreatment of any secure hash operates to chip a hash field for the RREQ packet and the projected approach looks terribly promising compared to other proposed solutions in literature.

Dhruva Patel et. al. [7] system describes various safety hazards which affect different levels of safety. Because of its simplicity and self-routing nature, MANET is responsive to various threats, and various attacks can breach separate layers within the network. There are numerous attacks and each attack has its own impact on different layers since some can only affect a single layer of the network while others can hit other layers, i.e. depending on the nature of the intrusion on how it reacts. Wormhole assault is now a network layer intrusion capable of crippling the communication network entirely. all the attacks.

Juhi Biswas et. al. [8] system present AODV With the implementation of the MANET and Wormhole Attack Detection and Prevention (WADP) technique on this revamped AODV, the routing protocol is modified to detect and prevent wormhole attacks in real world. So that fraudulent nodes are found within the network and the false positive limitation

node authentication function is eliminated. Simulation tests also indicate that node authentication does not eradicate false positives although it helps to chart the actual location of the hole jointly and could be a relatively double test for hole attack detection. This algorithmic rule doesn't use any special hardware for police work wormhole attack.

Kapil Raghuwanshi et. al. [9] take Initiative to eliminate or minimize the impact of a hole assault by offering a solution within the initial route setup point that could sense the presence of a void. Such resolution is based on the hop count analysis approach, i.e. hop count is used as a parameter for characteristic shapes involving tunnel space. Hop counts are evaluated to detect hostile nodes. Simulation of the planned work is completed in many node and traffic situations in the presence of a wormhole attack. The simulation results The predicted strategy demonstrates superior performance as PDR and attendance decreases however, in fact, "normal end-to-end gap" rises. It has been noticed within the analyzed state of affairs that the MAODV has a superior performance than AODV. Modified AODV is ideal for wormhole attack identification and bar detection. This strengthens the insecure circumstances of the packet distribution level, with a borderline reduction in turnout and a reasonable raise in end-to-end delay.

System [10] Throw light on a weak wormhole attack, during which two or more malicious nodes type a tunnel as a relay mechanism into the packets themselves. Such an attack may result in the packets sent being selectively diverted, fabricated and changed. During this paper AN identity is planned on the side of the clusters to shield the network from hole attacks based primarily on the signature theme. The designed theme need not transmit any credentials between nodes, thereby overhead calculations. Cluster During which cluster heads area unit selected in such a way that they cannot be harmful, specifically focused architecture is employed. This style works in 3 steps. The effects of the simulation indicate the improved performance of the theme in terms of throughput, packet delivery ratio and end-to-end delay.

According to [11] proposed protocol is based on Neighbourhood overhearing and frequent analysis of key and different tables and expected procedure information was found to be safe and a few region assault units are checked on it. Wormhole assault is observed by nodes being overheard. The results show that M-AODV has been strengthened in terms of the packet transmission magnitude relationship and the latency has therefore been lowered further, however the overhead amount has been raised. M-AODV jointly improves network reliability and stability. Thus, the planned protocol is proven to be protected in simulations alongside wormhole and blachole attacks.

Luo It. Al. [15] Fixes problems with WSN because every node is created to handle all the information from a forward-looking and network management-friendly, autonomous system to all levels, from the appliance layer to the appliance layer. This works very well, especially thanks to the algorithms developed with the short-shortcut WSN, they do not create simplicity and completeness when trying to implement widely and widely distributed different and low-energy WSNs at the same time.

Gante et al. [16] for example announce good organization of SDWSNs to boost potency and Overcome the inherent difficulties of common WSNs. The management theme is predicated on a Base Station Design for WSNs with Associated Integrated Controllers. The controller creates forwarding rules that area units stored in flow tables from location knowledge acquired through the localization technique processed in the layer layer application layer of the design.

Olivier et. al. [17] projected gradable design referred to as software package outlined Clustered detector It is believed that multiple base stations network (SDCSN) are used as controllers which additionally play the role of cluster head. Clusters of large nodes are divided into clusters, and each has a cluster head. The cluster head controls and coordinates the detector nodes in each cluster, and the knowledge processed in each cluster is diverted to the cluster head.

Oliveira et al. [18] style associated implement associate design supported Even within the framework many controllers in WSN referred to as Tiny-SDN supports Tiny-OS with a style structure consisting of SDN-enabled detector nodes and SDN controller node. This fixes issues such as in-band management, high communication delays and restricted power offers.

Tootoonchian et. a. OpenTM [19] projected associate very Put a little effort into this technique or work for new techniques observation SDN-based WSNs associated remains an exposed analysis space. Freshly but, work has been conferred on a network mensuration design supported SDN for observation of WSN data like routing path per-packet, the ratio each Link and hence the delay in each ointment-hop configuration.

East-West Apis modify supervisors that area unit within the similar dominion or neighbouring dominions to express with one another [20]. it's vital to say here that SDN isn't concerning SDNs have been shown to have the ability to change network management and allow innovation through network programmability, rather than enhancing network performance.

The challenges to think about once integration a quality management theme contains management

the impact of nodes coming In the network and therefore the nodes perform network functions, implement functions, and different network attributes on QoS. A number of solutions and process steps have been provided by Zhou et al to stop the problems associated with the SDN controller nodes and offer WSN efforts. [21].

III. PROPOSED SYSTEM DEISGN

Figure 1 describes the cluster creation process and transmission between source and destination node. At the end of each TS network nodes verifies sensed data and broadcast messages to nodes within given Cluster Distance (CD) for cluster creation. Cluster creation uses the Relay Node (RN) and CD to group the sensor in same cluster. Upon accepting the broadcasted message each node verifies the value of RN. If its value is within RN it stores in its memory and compares CD with each node's distance. If the distance between nodes is same or less to CD and sensed value is within given RN then those group of nodes forms a cluster. The nodes NID which are related they will not broadcast message for cluster creation. Nodes which are not participate in cluster creation process based on *RN* and *CD*.

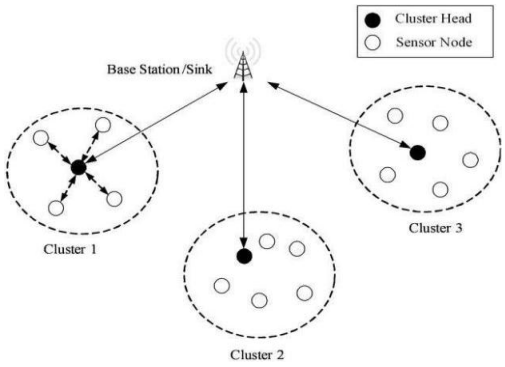


Figure 2: Proposed system cluster scenario

IV. CLUSTER HEAD SELECTION :

Input: Cluster set with nodes.
Output: Ch selection with remaining sensor node.
Step 1: select all nodes as initial population.
Step 2: Select evaluation set
Step 3: Apply crossover on similar power nodes.
Step 4: Apply mutation on each sensor node.
Step 5: Apply fitness on all nodes power
Step 6: select best node using rout let wheel selection.
Step 7: Check GA evaluations
Step 8: Select final max energy node as CH node.

V. DATA TRANSMISSION USING BTC :

A dynamic cluster it will be designed when the goal reaches several cluster boundaries. A difficult role problem is how the device identifies the situation, particularly in a fully distributed environment, when the goal reaches the boundaries. We use boundary nodes to solve this problem in a fully distributed way.

The following assumptions are taken in order to design the proposed algorithm.

In this work we have carried out an algorithm which is wormhole attack identification has been through in a cluster based network creation manner to eliminate the wormhole attacks. Basically AODV routing protocol is utilized as the fundamental network topology. A multi-layer method is used to determine if a node is participating in a wormhole attack. Introduces the layered approach to decrease the capacity of dispensation on respective cluster heads. Due to safety point of view, this also decreases the risk of a cluster head should be hacked by attacker.
The complete network is separated in clusters sets illustrated in figure 1. Sometimes clusters might be corresponded or separate. Every cluster contains a single cluster head as well as number of cluster member nodes. Member nodes forward on the data only to the cluster head when any nodes want to send data to Cluster Head (CH). The CH is responsible for forward on the collective data to all its other cluster members. The CH is selected enthusiastically and preserves the routing information.
Input: Primary source node *Sender_node*, Destination node *Dest_node*, Group of nearest nodes Neigh_node *[],* node id as *N_id,* node energy *N_eng;*
Output: From source to destination way based on the given algorithm.
Step 1: initially system select the *Sender_node* and *Dest_node* on dynamically
Step 2: select the packet or file f for info broadcast.
Step 3: if (*file or data* ! =null)
Step 4: read each byte *bytes* form *file or data* when reach null
Step 5: send data, initialize *cost_filed_1, cost_filed_2, parent_filed_1, parent_fileld_2*
Step 6: while (nd[i] when reach NULL)
cost_filed_1=node[i]_eng
parent_filed_1= node[i]_id
cost_filed_2 =node[i+1].eng
parent_filed_2= node[i+1] _id
Step 7: if (*cost_filed_1> cost_filed_2*)
cost_filed_2=null
parent_filed_2=null
Else
parent_filed_1= parent_filed_2
cost_filed_1= cost_filed_2;
parent_filed_2=null
cost_filed_2=null
Step 8: end of while loop
Step 9: reiteration till when extent at the sink node

VI. RESULTS AND DISCUSSION

After In this section we present experiment analysis using log file, Once simulation has finished it will create trace file at background, which contains all information of node communication as well as other log information. we have created a database of 5 text files which contains reading of 5 ms each till 25 ms as our simulation time is 25 ms. After that, we read the text file in a program created for the trace in Netbeans IDE 8.2. We got readings of various events in Netbeans from which we have plotted the graph of various parameters such as Drop rate(DR),Throughput and Packet Delivery Ratio(PDR) calculate d according to equation (1), (2) and (3) respectively. The evaluation has done with various WSN as well as cluster network existing protocol [12] [13] [14].

Table 1. The simulation parameters has used which is described in below table

Parameter	Values
Simulator	NS-allinone 2.35
Simulation time	25 sec
Channel Type	Wireless Channel
Propagation Model	Two Ray Ground
Standard	MAC/802.11
Simulation Size	1000 *1500
Max packet Length	1000
Ad hoc routing	AODV
Traffic	CBR

Table 2: shows the basic difference between the proposed and existing WSN.

Parameters	**WSN [21]**	**Proposed (cluster base with AODV)**
Data Aggregation	No	Yes
Data Security	Yes (selective)	Yes
Energy Conservation	No	Yes
Packet Loss	High	Low
End to End delay	High	Low
Packet Overhead	High	Low

1. Drop Rate:
It is defined as the number of packet lost per number of packet sent. The smallest amount value of drop rate states superior performance of the protocol.

$$Drop\ Rate = \sum_{i=0}^{n}\left(\frac{packet\ received\ [i \dots n]}{sent\ packet\ [i \dots n]}\right)$$

2. Throughput:
It is defined as the total number of packets supplied over the entire simulation era.
This is a mix of TCP's total number of packets and the total number of packets sent.
The higher throughput value means higher performance of the protocol.

$$Throughput = \left(\frac{\sum_{i=0}^{k} received\ packet[TCP]}{\sum_{i=0}^{l} sent\ packet[TCP]}\right) * 100$$

3. Packet Delivery Ratio (PDR):
The PDR is defined as the ratio of the data packet numbers to the number of packets produced through the network. The higher value of the packet distribution ratios represents greater efficiency of the protocol.

$$PDR = \sum_{i=0}^{n}\left(\frac{packet\ received\ [TCP]}{packet\ sent\ [TCP]}\right) * 100$$

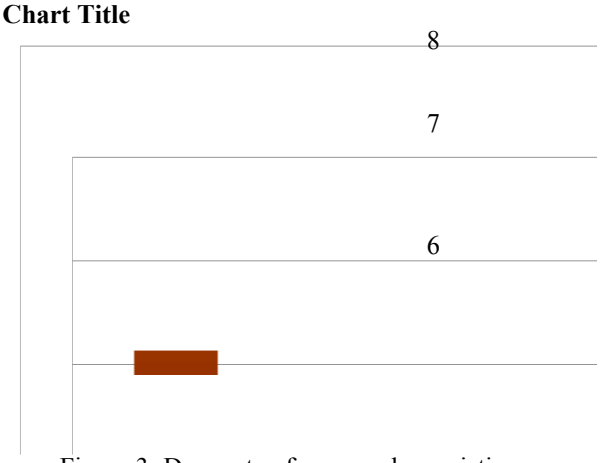


Figure 3: Drop rate of proposed vs existing

This figure 3 will provide the drop rate overall simulation during the communication with other protocols. The above graph has calculated based on various experiment analyses in NS2 environment. All defined protocols has used with different number of nodes in cluster network. The graph shows minimum packet drop rates of proposed AODV than other protocols.

Figure 4: Throughput of proposed vs existing
96
This figure 4 will provide the throughput of system during the communication with other protocols. Throughput s the most vital parameter the measure the QoS of any network. Similar as first experiment .tr files has95 utilized for evaluate the throughput for all protocols. This figure also shows our approach produces highest throughput than all three protocols.

Chart Title

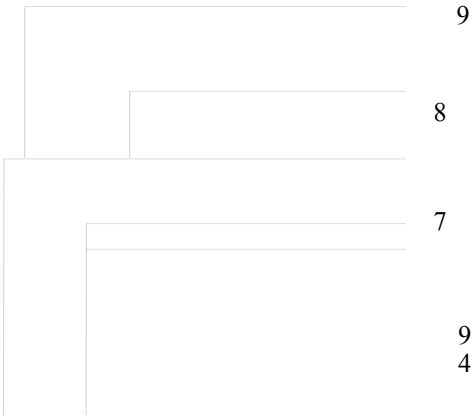


Figure 5: Network lifetime of proposed vs existing

This figure 5 will provide the how system will increased the actual time percentage of simulation due to proposed energy6 conservation protocol. Network lifetime has calculated based on after applying the energy conservation protocol, conservation s the technique during the data transmission eliminate unnecessary energy utilization by in Seconds

This proposes an aggregation of the HMAC and data in WSN. The CHs are initially chosen based on the connectivity of the node, which acts as a Data Aggregator. The proposed techniques$_3$provide the ad hoc network with greater security and also prevent different types of network attacks. It supports in the direction of escalations the packet delivery ratio (PDR) as well as decreases the network directly above by cultivating the enactment of the respective routing protocol. In the future, for faster detection of wormhole nodes, it will also need to modify the table entries in the receiver node. And also improved the security of the wireless ad hoc network. By means of set up such effective approach to prevent various kin fog network attacks with new defined algorithm. It also exposure and elimination of wormhole attack throughout data communication. The proposed Techniques offer greater security for various types of ad hoc networks and also deter network assaults. This facilitates the elevation of the packet distribution ratio (PDR) and reduces the overhead network by encouraging the enactment of the respective routing protocol. To get the detection of wormhole nodes faster, the table entries at the receiver node need to be improved in future plan as well. And the security of ad hoc wireless networks is also improved. By applying such an effective approach, the kin fog network will deter various attacks with new defined algorithm.

VII. CONCLUSION

sensor, receiver as well as internal nodes respectively. In proposed system BTC has
used for best path selection as well as energy conservation protocol.
Time

References

[1] Yashpalsinh Gohil, Sumegha Sakhreliya, Sumitra Menaria "A Review On: Detection and Prevention of Wormhole Attacks in MANET", International Journal of Scientific and Research Publications, Volume 3, Issue 2, February 2013

[2] Vikaskumar Upadhyay,Rajesh Shukla "An Assessment of Wormwormhole attack over Mobile Ad-Hoc Network as serious threats", Int. J. Advanced Networking and Applications Volume: 05 Issue: 01,2013

[3] Dhruvi Sharma, Vimal Kumar, Rakesh Kumar, " Prevention of Wormhole Attack Using Identity Based Signature Scheme in MANET " , Computational Intelligence in Data Mining. Volume 2. Volume 411 of the series Advances in Intelligent Systems and Computing pp 475-485. 10 December 2015, Springer.

[4] Elham Zamani and Mohammadreza Soltanaghaei, " The Improved Overhearing Backup AODV Protocol in MANET" , Journal of Computer Networks and Communications Volume 2016 (2016), Article ID 6463157, 8 pages. Hindawi

[5] Amit Kumar, Sayar Singh Shekhawat, " A Parameter Estimation Based Model for Wormwormhole Preventive Route Optimization" , IJCSMC, Vol. 4, Issue. 8, August 2015, pg.80 – 85 International Journal of Computer Science and Mobile Computing.

[6] https://en.wikipedia.org/wiki/Wireless_sensor_network

[7] Ashish Kumar Jain, Ravindra Verma, " Trust-Based Solution For Wormhole Attacks In Mobile Ad Hoc Networks" , Volume-4, Issue-12, November- 2015, Global Journal Of Multidisciplinary Studies

[8] Rajan Patel, Anal Patel, Nimisha Patel, " Defending Against Wormhole Attack in MANET" , Fifth International Conference on Communication Systems and Network Technologies, 2015 IEEE.

[9] Dhruva Patel, Parth Trivedi, M. B. Potdar, " A brief Analysis on Detection and Avoidance Techniques of Wormhole Attack in MANET" , Volume 117 - Number 16, International Journal of Computer Applications, 2015.

[10] Juhi Biswas, Ajay Gupta, Dayashankar Singh, " WADP: A wormhole attack detection and prevention technique in MANET using modified AODV routing protocol", 9th International Conference on Industrial and Information Systems (ICIIS), 2014. IEEE.

[11] Kapil Raghuwanshi, Amit Saxena, Manish Manoria, " An Enhanced Integrated Solution for Identification and Elimination of Wormhole Attack in MANET" ,Volume 110 - Number 7, International Journal of Computer Applications 2015.

[12] Uddin MA, Islam MM, Khanom R, Mosaddek M. Efficient and Secure Cluster based Routing Protocol for Wireless Sensor Network.

[13] Ouafaa I, Mustapha E, Salah-Ddine K, SAID EH. PERFORMANCE ANALYSIS OF SLEACH, LEACH AND DSDV PROTOCOLS FOR WIRELESS SENSOR NETWORKS (WSN). Journal of Theoretical & Applied Information Technology. 2016 Dec 31;94(2).

[14] Olivier, F.; Carlos, G.; Florent, N. SDN Based Architecture for Clustered WSN. In Proceedings of the 2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), Blumenau, Brazil, 8–10 July 2015; pp. 342–347.

[15] Upadhyay R, Bhatt UR, Tripathi H. DDOS attack aware DSR routing protocol in WSN. Procedia Computer Science. 2016 Mar 1;78(C):68-74.

[16] Hu, F.; Hao, Q.; Bao, K. A Survey on Software-Defined Network and OpenFlow: From Concept to Implementation. IEEE Commun. Surv. Tutor. 2014, 16, 2181–2206.

[17] Gante, A.D.; Aslan, M.; Matrawy, A. Smart wireless sensor network management based on software-defined networking. In Proceedings of the 2014 27th Biennial Symposium on Communications (QBSC), Kingston, ON, Canada, 1–4 June 2014; pp. 71–75.

[18] De Oliveira, B.T.; Margi, C.B.; Gabriel, L.B. TinySDN: Enabling multiple controllers for software-defined wireless sensor networks. IEEE Lat. Am. Trans. 2015, 13, 3690–3696.

A modified EDDKA routing approach to enhance Quality of Service (QoS) Enabled Data Dissemination in VANET's.

Mr Amol S. Dudhe[1] and Dr Salim Y. Amdani[2]
Department of Information Technology,BNCOE,Pusad, 445204,India
Department of Computer Science and Engg, BNCOE, Pusad, 445204,India.
amol_dudhe@rediffmail,com, Salim amdani@rediffmail.com

Abstract. **VANETs have gained great focus among the researchers since last few decades. The areas of great interest are the types of routing, and quality of service (Qos). The main challenge is to find the ways to counter the continuous-ly changing VANETs topologies and its high speed nature, and then determin-ing which routing protocols are best suited for a particular transmission type; and which ones provide more consistent and stable routing performances. Rout-ing in networking is one of the key factors in determining how effective and ef-ficient data packets can be transmitted within a network environment. As the topology is consistently ever-changing a reliable communication with needed QoS is of utmost importance. QoS with respect to routing protocol is a domain which demands improvement. Great efforts have and are still being made to improve the routing capabilities of the protocols in existence to en-hance QoS mechanisms in routing protocols. The simulation results indicate that the proposed approach can lead to improvements in terms of QoS metrics like Throughput, Packet Delivery Ratio, end to end delay, Energy Consumption & Residual energy.**

Keywords: VANETs, Qos, Throughput, Packet Delivery Ratio, end to end de-lay, Energy Consumption & Residual energy.

I. INTRODUCTION

Traffic safety is a major challenge recognized by the major players in the automo-tive industry and by many governments. According to which annually thousands of road accidents are reported in any country. Traffic accidents are most of the day's results of the driver's failure to access quickly and properly the driving condi-tions. Normally drivers have imperfect information about road situations, speed and position of vehicles around them and typically are compelled to form decisions like breaking and lane changing without the benefit of whole data. "The need for commu-nication when the deployment of any fixed infrastructure is impossible and therefore the advancement of computer and wireless communication technologies, led to the event of Mobile Ad-hoc Networks (MANETs)" [1]. Vehicular Ad-hoc Networks as a subset of Mobile Ad-hoc Networks which provide data exchange via Vehicle-to-Vehicle (V2V), Vehicle to Roadsides (V2R) and Vehicle to Infrastructure (V2I) communications and a car which takes part in such a network is equipped with a WLAN and cellular communication device [2]. VANETs is additionally defined as a wireless communication technology which is additionally ready to enhance driving safety and velocity by exchanging real-time transportation information, and "it should upon implementation, collect and distribute safety information to massively re-duce the amount of accidents by warning drivers about the danger before they real-ly face it" [3]. In addition, VANETs also are ready to minimize incidents and improve traffic conditions by providing vehicles, drivers and passengers with information about the road condition. VANET has its own unique characteristics in compari-son with other sorts of MANETs, the unique characteristics of VANET include: pre-dictable mobility, lack of powers constraints, variable network density, Rapid changes in network topology, High computational ability and large scale networking [4]. Safety services information such as traffic accidents and road congestion which are sensitive to reliable and real-time communication should be broadcasted immedi-ately. Data transmission in such environment is critical and has got to be distributed in multiple paths to enhance the end-to-end delay. Some stale routes are generated in the routing table which leads to unnecessary routing overhead causing frequent link fail-ures as well as route discoveries. Therefore the discovered route between couple of vehicles should be as stable as possible to satisfy QoS requirements [5]. The intermit-tent nature and short-live of these algorithms, make the created clusters to

provide scalability with lower communication overhead [5]. Rapid change in topology, owing to time varying vehicle densities and other factors; both external and internal, makes preserving a route very difficult and this in turn incurs high routing overheads as well as low throughput [6]. A cluster on demand minimum spanning tree with primsalgo-rithm has been proposed in [7] for VANET. In this approach the vehicles has been clustered by accounting the intra-cluster QoS. An extended Kruskal algorithm has been proposed in [8] to support QoS.

Thus, we aims to develop a QoS Enabled Data Dissemination using Kruskal's algo-rithm to provide efficient data dissemination (QoS-EDDKA) and quality of service in VANET. This approach constructs the clusters aid of minimum spanning trees in every road segments by considering the intra cluster QoS. Each spanning tree will have a cluster head that is responsible to collects or disseminates the data from the leaf nodes and to other coordinator nodes and vice versa

II. RELATED WORK

A bee communications inspired QoS routing scheme to improve the network throughput and to reduce packet loss in vehicular ad hoc network (VANET) environ-ment has been proposed in [9]. Their scheme was inspired by the natural bee commu-nications paradigm. The routing scheme is inspired by the bee's foraging behavior in the nature. Control messages are exchanged within all nodes in the cluster to ensure a reliable and stable connection. However this may contributes to overhead problem in routing.

In [10], the authors present Cross layer based vehicular routing model (CL-RVR), to facilitate reliable routing in VANETs, which would cater to QoS requirements for desired applications by combining the parameters from physical and network layers. Their simulation results show that their proposed scheme can improve the packet delivery rate and reduce link failure ratio in routing in VANETs

A cluster-based routing protocol for urban Vehicular Ad Hoc Networks (VANETs) using Optimized Link State Routing (OLSR) was presented in [11]. In their model OLSR, MultiPoint Relays (MPRs) used for routing, are selected at each node using neighbors' reachability resulting in a high percentage of MPRs.

In [12], the authors present a link prediction algorithm as an addition to the AQRV protocol that selects the best routing path in urban vehicle environments to predict the duration of availability of the present path. This methodology expects to anticipate a connection breakdown before its incidence and directing the data packets through a substitute route.A QoS Enabled Data Dissemination using Kruskal's algorithm to provide efficient data dissemination and quality of service in Hierarchical VANET [13].

III. SYSTEM MODEL

The proposed system model consists of n number of vehicles, Road Side Unit (RSU). Every vehicle can communicate with other vehicles and also with RSU. QoS provisioning often needs negotiation between host and network, call admission con-trol, resource reservation, and priority scheduling of packets [14]. A communication link between the nodes will be exist if they present in their communication range The vehicles present within the radius of the intersection point are not allowed to form the clusters. Figure 1 shows the Clustering of Vehicles in VANETs with the proposed technique. The vehicles which are away from that region can form and join the clus-ters. It has been assumed that if the road contains any of the Road side unit then it will act as the cluster head for that region. Vehicle clustering represents a management scheme in which the adjacent vehicles are gathered into a group known as a cluster. Each cluster has one active node plays the role of a CH. CH is selected to control and manage the cluster activities. All the other nodes in each cluster are entitled cluster members. These members are usually belonging to one cluster or in some cases it belonged to multiple clusters. The cluster head (CH) can communicate with the clus-ter member (CM) as well as with the other cluster heads.

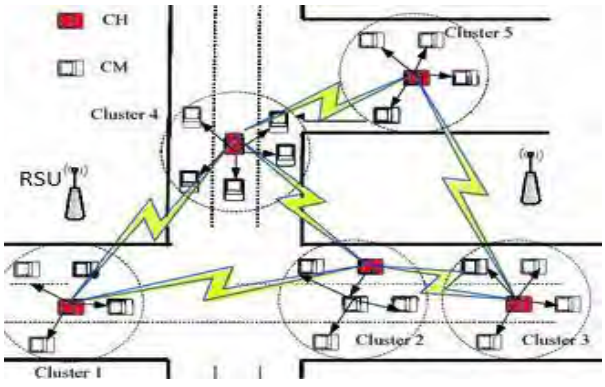


Fig. 1. Clustering of Vehicles in VANETs.

The number of clusters can be formed on a road segment can be determined by us-ing following equation Here l is the length of the road segment measured in meters; TR is the Inter-cluster communication range of the nodes. According to this the Road has been segmented and the cluster head has been selected in each segmented road. Figure 2 shows the segmentation of road in the

proposed approach. The clusters will be formed for the nodes which are travelling on the same direction.

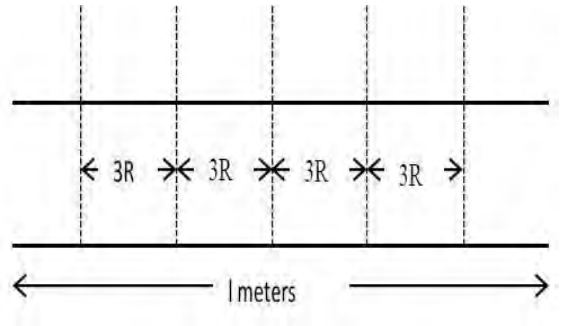


Fig. 2. Segmentation of Road.

The node stability has been computed based on the Time to Live (TTL) and the ac-celeration of the vehicle

Where D is the distance between the node and the road segment, S is the speed of the node.The stability of the node can be given as follows

Where size of the queue is the present size of queue at node i, length of the queue is the queue length between the node i and j.

The probability density function (pdf) has been calculated for every node relied upon the NS value.

$$\frac{}{\sqrt{=}} \quad (\ -)[\text{—}]$$

Here μ is the mean of the NS value gotten from the nodes in the road segment, σ is the standard deviation valued to the particular mean.

The signal to interference ratio between a node i and node j can be given as follows

Vehicle pseudo ID	Temporary public/private key pair	Pairwise key

$$\frac{}{\sum}$$

Here, N is the number of channels available between the node I and node j, ptx is the transmitted power between the two nodes i and j. Prj,k is the measured received power at the node j on the channel k.

The weight for each link based on NS and can be given as follows

Initially the VANET Topology was created by SUMO 30.0.0 & MOVE.jar in the Ubuntu 14.04 LTS. The vehicle (CM) which wants to send the message communicat-ed to the corresponding cluster head for transmitting the message by sending group information request. In order to provide data security within the message digital sig-nature was attached by using ECDSA algorithm accordingly it generates the Private and public key and also it encrypt the data. The road side unit (RSU) holds both Pri-vate and public key information. Every vehicle which send group information request message was encrypted by private key of his own with certificate attached as shown in table 1.

Vehicle ID	Message ID	Message	Timestamp	Certificate

Table 1. Frame format for group information request messagThis message is signed by public key of RSU and it forward to trusted authority (TA). TA decrypt this message and check certificate of vehicle from its database. If its valid then it sends verification message to RSU. TA assigns Pseudo ID to vehicles and temporary Public and Private Key pair. TA sends this Pseudo ID, temporary pub-lic and private key pair this message is encrypted by permanent public key of vehicle as shown in table 2.

Vehicle ID	Pseudo ID	Temporary public/private key pair	Timestamp

Table 2. Trusted authority table.

In this way all the vehicles gets authenticated and RSU form cluster of vehicles and also it select group leader (GL) or cluster head (CH) and assign group ID .The format of it is as shown in table 3.

Vehicle pseudo ID	Pseudo ID	Temporary public/private key pair	GL/GM

Table 3. RSU vehicle information table

Cluster head (CH) or group leader (GL) maintains table of vehicles in group and which has vehicles Pseudo ID and their Temporary Public and Private Key pair as shown in table 4.Cluster head then Broadcast this message to all vehicles encrypted by group key.

Table 4. Cluster head table

IV. ALGORITHMIC STEPS

Start:

Algorithm packet Scheduler (Graph T){

Set priority queue Q which composed of all the edges of G

Remove all the Successive tree links for node(Vertex V) to first trust node. R <- Total no of removes links

Best candidate <- b

Run BFS Algorithm from v in a radius

// Problem description : Hierarchical Vanet Packet Scheduler For every vertex V in V[G]:

If nodes other than b on disconnected main tree: Candidates <- nearest Candidate nodes into tree list Best weight =infinity

For each node g in the candidate list: Hg <- Hop count from g to the sink Dg <- Degree of node g

If (Dg+Hg)< Best weight:
Best weight=Dg+Hg
Best candidate <-g
Connect node V to Best candidate in shortest path.

In our work we Run BFS Algorithm from v in a radius to from the minimum span-ning tree in which best weight considered as infinity, nearest candidate with minimum weight is added to tree list. For each node g in the candidate list Hop count (Hg) from g to the sink and Degree of node g (Dg) is calculated .If sum of Hop count (Hg) and Degree of node g (Dg)is less than then the Best weight is modified as Dg+Hg and node g is considered as best candidate. Connect node V to Best candidate in shortest path.

V. SIMULATION RESULTS

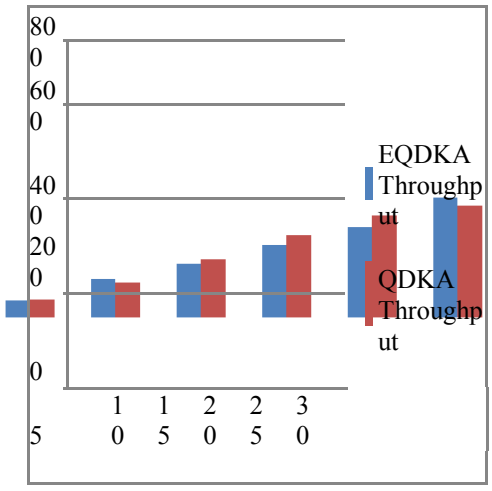
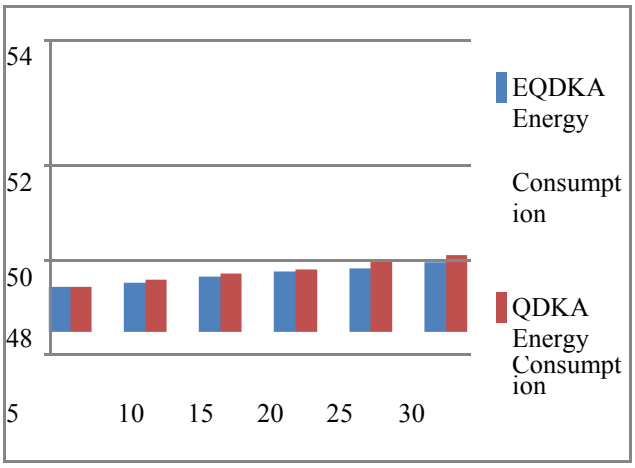


Fig. 3. Comparative Analysis of QDKA & Proposed EQDKA algorithm on Throughput with respect to time



ig. 4. Comparative Analysis of QDKA & Proposed EQDKA algorithm on Packet Delivery Ratio with respect to time

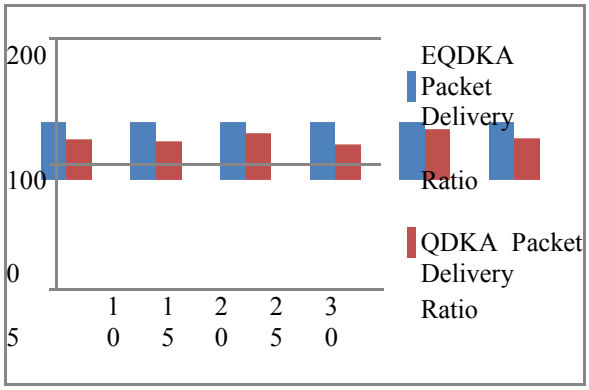


Fig. 5. Comparative Analysis of QDKA & Proposed EQDKA algorithm on end to end delay with respect to time.

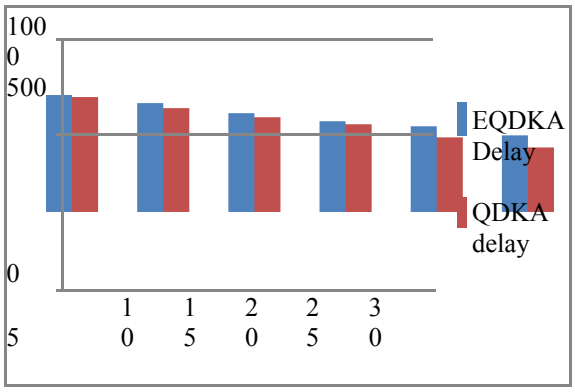


Fig. 6. Comparative Analysis of QDKA & Proposed EQDKA algorithm on Energy Consumption with respect to time.

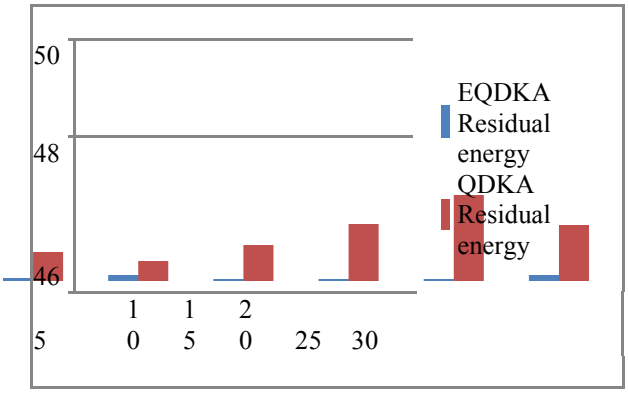


Fig. 7. Comparative Analysis of QDKA & Proposed EQDKA algorithm on Residual energy with respect to time.

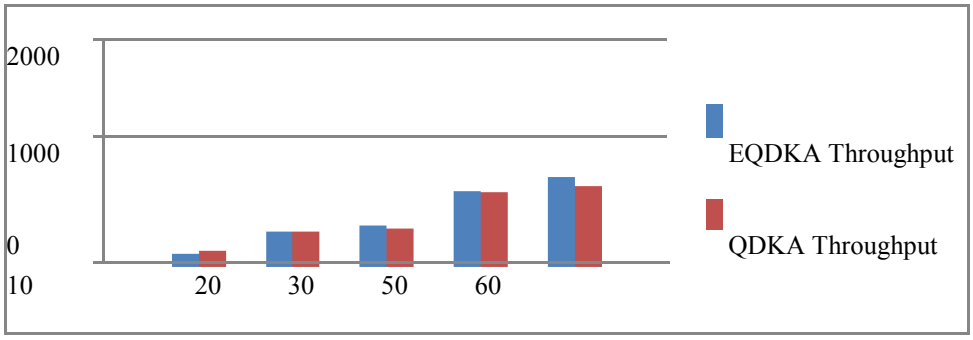


Fig. 8. Comparative Analysis of QDKA & Proposed EQDKA algorithm on Throughput with respect to number of nodes.

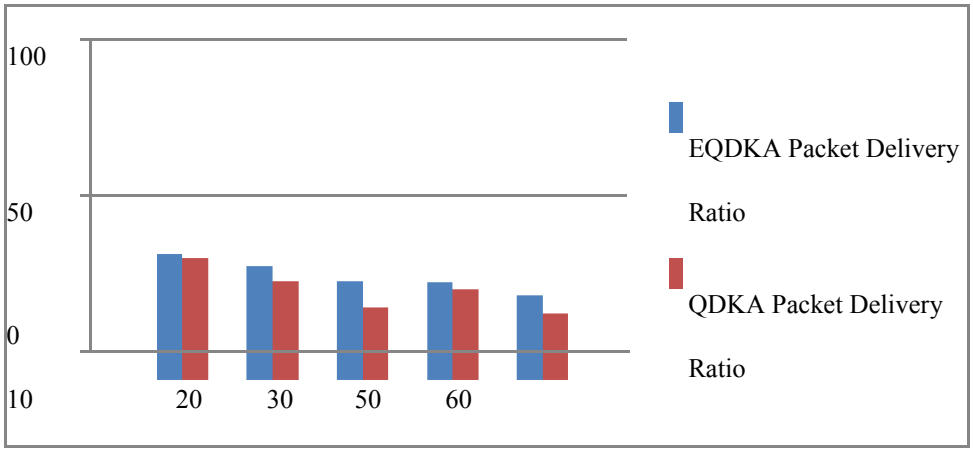


Fig. 9. Comparative Analysis of QDKA & Proposed EQDKA algorithm on Packet Delivery Ratio with respect to number of nodes

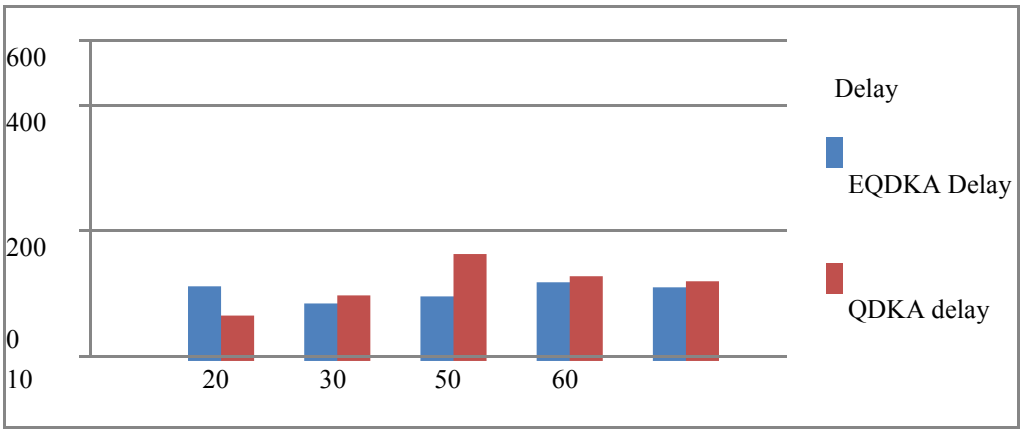


Fig. 10. Comparative Analysis of QDKA & Proposed EQDKA algorithm on end to end delay with respect to number of nodes.

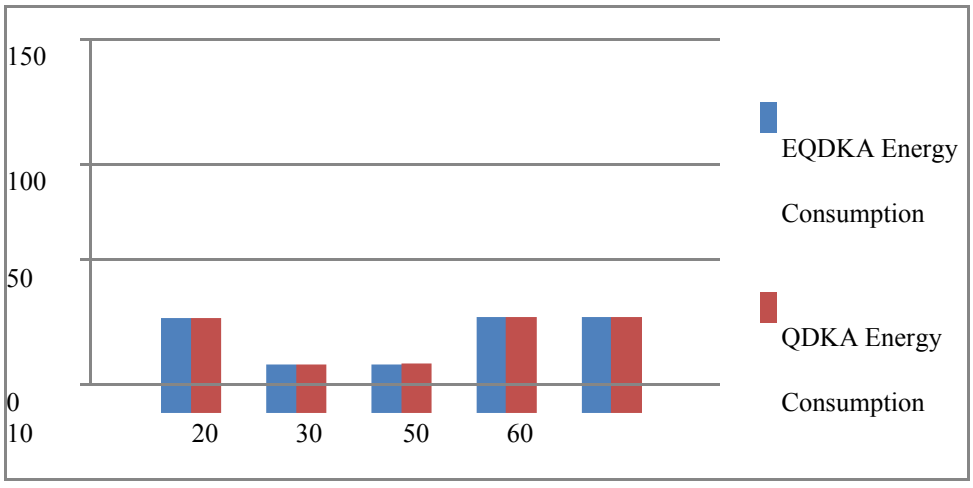


Fig. 11. Comparative Analysis of QDKA & Proposed EQDKA algorithm on Energy Consumption with respect to number of nodes.

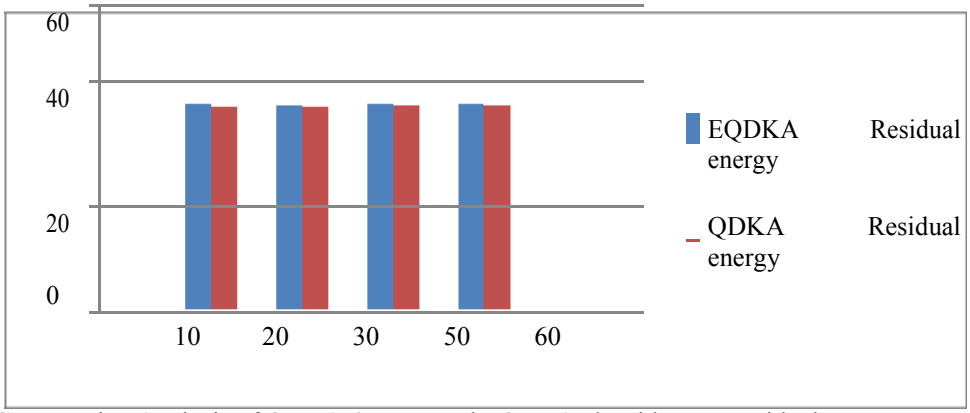


Fig. 12. Comparative Analysis of QDKA & Proposed EQDKA algorithm on Residual energy with respect to number of nodes.

VI. CONCULSION

Through the stimulation results obtained from NS2 as shown in fig. 3 to fig. 12 we can conclude that by using proposed EQDKA algorithm Qos have been improved with respect to Throughput, Packet Delivery Ratio, end to end delay, Energy Con-sumption & Residual energy.

References

1. Khairi. D, Berqia. A,"Survey on QoS and Security in Vehicular Ad-hoc Networks.", In: International Journal of Advanced Research in Computer Science and Software Engineer-ing, vol. 5, no.5, pp. 42-52.(2015).
2. Shrivistava. P, Ashai. S, Jaroli. A, Gohil," Vehicle to Road-Side-Unit Communication Us-ing Wimax," In: International Journal of Engineering Research and Applications, vol.2, no.4, pp. 1653-1655. (2012).
3. Adibi. S, Erfani. S,"Mobile Ad-hoc Networks with QoS and RSVP Provisioning," In: CCECE,Saskatoon, Canada, (2015).
4. Al-Sultan. S, Al-Doori. M, Al-Bayatti. A, Zedan.H " A comprehensive survey on vehicular Ad Hoc network," In: Journal of Network and Computer Applications, vol. 37, pp. 380-392 . (2014).
5. Abubakar Aminu Mu'azu and Sallam Osman. Fageeri , "Enhanced Bandwidth Reservation Guarantees for QoS Routing in Vehicular Network"in 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), Khartoum Sudan, pp.12-17.(2017).
6. J. Bernsen and D. Manivannan, "Unicast routing protocols for vehicular ad hoc networks: A critical comparison and classification," In: Pervasive and Mobile Computing, vol. 5, pp. 1-18, (2009).
7. Kponyo J.J, Yujun Kuang, Enzhan Zhang, Domenic K, "VANET Cluster-on-Demand Minimum Spanning Tree (MST) Prim clustering algorithm" In: International Conference on Computational Problem-solving (ICCP), Jiuzhai, pp. 101-104.(2013).
8. Mohamed Aissa, Adel Ben Mnaouer, Rion Murray, and Tobago Abdelfettah Belghith, "New Strategies and Extensions in Kruskal's Algorithm in Multicast Routing" In: Interna-tional Journal of Business Data Communications and Networking, vol. 7, no. 4, pp. 32-51. (2011).
9. Low Tang Jung,et al," Bee Inspired QoS Routing in VANET " In: 2014 4th World Con-gress on Information and Communication Technologies (WICT 2014), Bandar Hilir, Ma-laysia, pp. 176-181. (2014).
10. Mahadev A Gawas,et al," Cross Layer Approach for Neighbor Node Selection in VANET Routing "In 2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP), IEEE, pp. 121-127. (2018).
11. Maha Kadadha,et al," A Cluster-Based QoS-OLSR Protocol for Urban Vehicular Ad Hoc Networks ", in 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), Limassol, Cyprus, pp. 554-559. (2018)
12. M. Naresh,, Aayushi Raje, Karanam Varsha, " Link Prediction Algorithm for Efficient Routing in VANETs " In: Proceedings of the Third International Conference on Compu-ting Methodologies and Communication (ICCMC 2019), PP 1156-1161.(2019).
13. Dr.B.Mukunthan,et al," A Novel Krushkal's Algorithm Approach to Ensure Quality of Service Enabled Data Dissemination in Hierarchical Vehicular Ad-Hoc Network" In: IRACST – International Journal of Computer Networks and Wireless Communications (IJCNWC), Vol.7, No 4, pp. 6-12.(2017).
14. Saharan. S, Kumar. R "QoS Provisioning in VANETs Using Mobile Agent," In: Interna-tional Journal of Computer Science and Communication, vol. 1, no. 1, pp. 199-202.(2015)

A Comprehensive Study and Possible Solutions on Hostel Management System

Lakshmi K. Jha, Harshali P. Patil
Department of Computer Engineering
Thakur College of Engineering and Technology,
Mumbai University
lakshmijha19@gmail.com, harshali.patil9@gmail.com

Abstract—**The ongoing manual hostel management system demands a lot of paper work and calculation & therefore may be imprecise, which leads to inconsistency and inaccuracy in maintenance of data. The information stored on the paper can be lost, stolen or destroyed, under any circumstances. Thus, the presentsystemtakeslotoftimecausinginconveniencetohostel ler and employees. Due to this manual behaviour, it become complicated to add, update, delete or view the data. Also, if number of hostellers increases drastically, then to preserve detailed record of every student is extremely difficult. Therefore, thekeyintentionofthispaperistargetedforCollegeHostelwhich integrates allocated room report, unallocated room report, transaction management, maintaining the decorum of the hosteller&giveplatformforplacementstobeexcellentlycontrol. The paper proposed management of hostel in computerized system. The system is equipped with some special features for serving hostel admin. Each record will generate a unique identity and it can be searched by a unique ID. The warden will have a bettercontroloverthefunctionalitieslikecomplaintmanagement and tracking time using RFID. This will turn down efforts of warden & provide better service to thehostellers.**

Keywords— Hostel management, Active RFID, complaints management, Online hostel system, Hostel placements.

I. INTRODUCTION

21st century is the century of knowledge. Educational institutions and hostels are the two sides of a coin. Thus, the number of hostels is also widely increasing for providing the home like atmosphere to accommodate the students. And real time chaos occurs for the persons who are searching for the hostels and software or online applications are not usuallycast offinthiscontext.Furtherthepersonswhoreallywantstoserve at hostels places finding difficulty to get jobs. This paper proposesthesystemwhichwillovercometheproblems likeAllot different students to the different hostels. -Vacate the students for the hostels. -Control the status of the feepayment.
-Manage the details of the students & amend the student records. Hostel infrastructure Mess creation and management RegisteredstudentsMeritlistsRoomandmessallotmentMess billcalculationFinesandPayrollandgivingopportunityforjob seeker for laundry, rector, electrician, carpenter, Messmanger, etc.

II. LITERATURESURVEY

"Tracking Student Movement Single Active RFID" illustrated a model for tracking movements of students by using SMOSA which comprises of two graphical user interface (GUI) that standalone,onlineuseranddatastorage;database.Thissystem supervised two tasks: attendance system and tracking system by use of RFID[1]-[5].

In 2017, "Solving Hostel Student Issues Using Mobile Application"waproposedtosolvethehostelstudentproblem like food, water, security, & electricity issues in digital environment. In this system, one can send the message to the hostel management via mail or whatsapp messages and these messages will substantiate by hostel management. The paper depicts web-based application is better to use instead of using paper [6],[7].

Indocon micro engineers limited established a hostel management software having distinct features such as modifyingtechnology,income/revenuemanagement,cohesive web booking engine and interfaces to all engagement channel [8].

For further examine, we examined the College Hostel Management Software developed by Initio which covers six modules like the hostel module, the transport module, the librarymodule,theregister/storemodule,thequerym oduleandthe visitors tracking module. It bids information on the building, rooms and students [9],[10].

ThepaperbyC.Prom&Rosyproposesasystemwhichin dulge hotelier with mobile technology era. It shows an android application named ALICE which is divided into categories of SPA,housekeeping,front-desk,concierge,roomservice,valet- travel & TV & appliance. The aim of the system is to create fluid 2-way communication with the hotelier to improved hospitality. The hotelier can directly communicate with this android app to control the requirement like controlling the temperature of a/c, requesting for room service. This systemis programmedtoactonpredefinedthings.Byselectingas ervice, an alert message will be sent to management, where further action is taken[11],[12].

Thepaperby'MaryShalin'proposesanandroidapplica tionfor hostel Out pass form, which allow a student to be able to fill form in an elegant way rather than doing it manually. The application makes generation of a completed form to get an Out-pass form in a simplest way. This reduces a lot of human efforts and also use of paper to a small-extent. The paper aims to give effective result in case of emergency situations, when the student might not be able to contact parents or wait for an approval due to several reasons. This will result in less time consumption&aneasywaytokeeptrackofoutpassreco rdof student with mentioned time[13]-[15].

III. PROPOSEDSYSTEM

This paper defines managing a hostel. The research work described in this write-up is marginals to educational institutions. The aim is to provide solution of the problems of traditional method of managing hostel facility. The systems appeal to attempt hostel services related to administrator, management and student of the hostel. It self-operates the administrative processes and turn down the stress associated withhuntingforinformationonastudentorafacilityina stack ofregisters.Itispeculiarlyprogrammetocentrallyalloc ateand maintain accommodation spaces in a typical student's hostel. Theuserofthesystemcanupdatethedataofallthosestud ents who have left the hostel, and check the profile of newly come on the tip of finger. The student staying in the hostel will be registered with an ID number allocated at the time of room allocation. The generation of the bills is automated and is directly issue in the form of notification to thestudent.

The system has a unique approach to trace information andanother perspective of the system is for job hunting process. This feature will provide a platform for workers to seek jobs. The module will be categorized as laundry men, rector, mess staff&in-charge,electricianandacarpenter.Thiswillleadthe system to follow up the efficiency, which will automatically decreasetheunemploymentbyprovidinganareatoactu pon.

Nextaspectofthepaperhighlightsthedigitaleraofinno vating an online application by trim back the paper. This will bring a solutionfortheissuefacedbythestudentsstayinginthe hostel.
Issuecanbemessproblem,electricityproblem,waterp roblem, food problem, furniture problem, security problem, medical service problem & internet facility problem. To handle the mention issues, this module will invent an automated application to fire complaint, give suggestions and feedbackto thepanel.

LaststageincorporatesuseofRFIDtags.Thelightisthr ownin trackingtheincomingandoutgoingactivitiesusingthe tagand receiver. It will enhance security and record is maintained. In addition, it drives an even-flow operation in accessing and retrieving data in a polished way and hence resulting in saving a lot of human effort and time.

Table I: Functionality of modules

Sr. No	Modules	Modules Functionality	Processing
1.	Registration	Entering details of hosteller.	User profile.

2.	Room Allotment	Authenticate a user & allotting room.	Hosteller details with Registration number& Roomno.
3.	Fee Management	Fee details accesses by administrator.	Fetching data of payment & generating Receipt.
4.	Complaint Management	Offline text message of units.	Complaint ID & Actions taken.
5.	RFID Tracking	Senses incoming and outgoing time.	Communicates with the tag & receives Information.
6.	Offline Notification	Providing timely information to parents	Alert message.
7.	Job Hunting	Job related to hostel	Checking availability of workers.

The work of each module is as follows:report generation which is decisive in data handling, which makeproposedsystemmorereliable,morerobustandefficient. Thisapproachwillcoordinateusingacentraldatabasetohandle the complexness of student's hostel management and administrativefunctions.*Registration:* Once the hosteller installs the application, then they would be asked to register themselveswiththeirdetails.Thiswillbestoredinthe database and a unique serial number isgenerated.*Room Allotment:* This section provides an online criterion for the admission to hostels. Hostel accommodation can be provided based on the availability of rooms. The administrator will preserve the records of room vacancies & wipe out particular record from the datastore. Details like number of rooms in each hostel, number of members accommodate in eachroom.

A. *Fee Management:* The system handles the monotonous task calculating hostel fees in easierway inhaling faster access. For every accommodation, software allows administrators entrust with fare that hosteller needs to pay and gives them payroll for the same.

B. *Complaint Management:* The module will work with anapproachtomakecomplaintorsomesuggestionor feedbacks to be easier to coordinate, monitor, track and resolve issues put on by the hosteller in order to increase the performance of the management & aiming for bestimprovements.

C. *RFID Tracking:* Hosteller incoming and outgoing time will be recorded using RFID smart labels. When thehostellerIDcardwillbeswipedagainstthereader, information will be matched in existing system of the datastore & time will bemarked.

D. *Offline Notification:* An alert message of the above module will make acount.

E. *Job Hunting:* Job seekers will have an opportunity to search jobs in this platform. They will be able to identify job, analyse requirement & apply accordingly. The background process will check availability of workers & notify them for thesame.

The flow of this system is as follows:

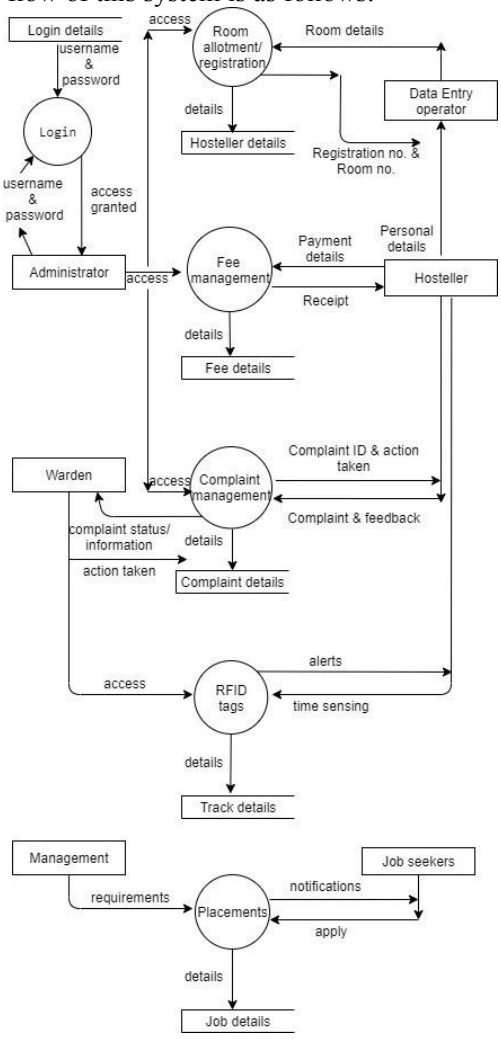


Fig. 1 Flow of proposed system

IV. CONCLUSION

Thepaperproposedauser-friendlyapproachforstudentHostel. The paper roots

on the requirements specifications of the hosteller with the analysis of the previous system, scope to the flexibilityof enhancement in future work. The system will ease all the members using this approach, to manage and automate overall processing of any largehostel.The idea highlights on managing hostel rooms, hosteller records,&roomallocationprocessetc.Theschemeap proaches flexible recommendation and can be reform according to the individualneeds.

V. References

[1] Herdawatie Abdul Kadir, Mohd Helmy ABD Wahab, Zarina Tukiran, Ariffin Abdul Mutalib, "*Tracking Student Movement using Active RFID*", International Conference on Applications of ComputerEngineering.

[2] ChristophJechlitschek,"*ASurveyPaperonRadioFrequencyIdentification (RFID)*", IJCSE, Vol 1, Issue 3, Nov 2013.

[3] Farooque Azam, Iqra Basharat, "*Database Security and Encryption: A Survey Study*", International Journal of Computer Applications (0975 – 888) Volume 47 – No. 12, June2012.

[4] Khan, S.A., Qaiser, A.; "*Automation of Time and Attendance using RFID Systems"*, International Conference on Emerging Technologies, ICET '06, 2006, pp.60-63.

[5] Roy Want, "*An Introduction to RFID Technology*", 1536-1268/06, 2006 IEEE.

[6] K. Mehaboob Subhani, Bhupendra Singh, Manoj Kumar D S, "*Solving Hostel Student issues using Mobile Application*", 2017 International Journal of Pure and Applied Mathematics, ISSN:1314-3395.

[7] O.Duduyemi,andO.Shoewu;S.A.Braimah,"DesignandImplementation of Hostel Management System (HOMASY)": LASU as CaseStudy"

[8] "*Indocon Hostel Management Software*",Indocon Micro Engineers Limited, pp: 23 Nov, 2016, NewDelhi."

[9] Emmanuel Adu Baffoe,, "*College Hostel Management Software*", June 2015, Initio.

[10] Kola Ayanlowo, O. Shoewu, Segun O. Olatinwo, Olusegun O. Omitola, Damilola D. Babalola, "*Development of an Automated Hostel Facility Management System*". Journal of Science and Engineering. 5 (1):1-10.

[11] C.Prom&Rosy,"*AstudyonHostelReservationTrendsofmobileapp via smartphone*", IOSR Journal of Computer Engineering (IOSR-JCE),pp:01-08

[12] Bikash Choudhary, Deepak kumar, Deepika Priyadarshani Khatua, Ajit kumar Patro, "*Online Hostel Management System*"; International Journal of Engineering & Technology, ISSN: 2394 – 627X, Vol. 4, No. 3,(March,2017).

[13] Mary Shalin Benigna, *"Android App for Hostel Outpass Form*", International Journal of advanced research, ideas and innovations in Technology, ISSN:2454-132X.

[14] Jianye Liu, Jianka Yu, "*Research on Development of Android Applications*" in Fourth International Conference on Intelligent System-2011

[15] Jadhav Snehal Balasaheb, Supekar Bhagyashri Sitaram, WakodeVrushali Khushalrao,VasaikarNikitaAshok,MandlikPriyankaBhausaheb"*Web Based College Admission System*" in IJEDR-2014

Multidevice Farm Monitoring Cloud Network and Smart Crop Production Enhancement System

Siddhesh Vartak,Dr.Harshali Patil
Master of Computer Enginnering
Thakur College of Engineering and Technology
Mumbai,India
vsid55@yahoo.co.in

***Abstract*— the combination of present day data technologies into modern farming has just added to yieldincrementsinthemostrecentdecades. These days,thedevelopingInternetof Things(IoT) alongside WirelessSensor Networks(WSNs)with their minimal effort sensors and entertainers empowernovelapplicationsandnewopendoors for aprogressivelyexact,webpageexplicit,and supportable agribusiness with regards to Smart Farming. Right now, present an all-encompassinghorticultural checking framework, its structure, anditsbuildingusage. Theframeworkbasically centers around in-situ appraisal of the leaf territory record(LAI),asignificantharvestparameter.Also, we present true difficulties and encounters picked upindifferentarrangements. Atlast,first outcomes are exemplarily exhibited so as to quickly address the capability of ourframework.**

KEYWORDS: IOT- INTERNETOFTHINGS

I. INTRODUCTION

Farm Monitoring using cloud network aims to increase overall development of agriculture filed. It aims to replace the traditional techniques with new wireless sensor technology. Farm Monitoring is a system that connects digital technology with traditional systems of agriculture. Thissystemwillenhancethe communication and sharing experiences between farmers and experts and others interested in this technology.Thissystemisdesignedtodevelopasmart farming system that uses advantages of cutting edge technologies such as IoT, wireless sensor network and cloud computing to help farmers to enhance the way of farming is done. Using [1] sensors like temperature, humidity, moisture etc. are used to collect information about the field and help farmers to take precise decisions on insights and recommendations based on thecollecteddata. Agricultureisoneofthemost important sectors of Indian Economy. Indian agricultural sector accounts for 18 % of India's gross domesticproduct(GDP)andprovidesemploymentto50% ofthecountrys' workforce.EstimationofCrop Yield during the cropping season plays an important role for planners and policymakers for decision making. It is one of the critical parts as it depends on different factors weather, soil and crop Management. Weather plays one of the important role for plant growth and development, soil is important for plants to get nutrients while crop management is important for planning the planting time and application of different management practices for better crop yield. Different approaches had been reported for estimating crop growth and development based on statistical and mathematical based models. In this study the authors have tried decision tree and random forest based Machine Learning approaches to estimate crop yield. DecisionSupportSystemfor Agro-technology Transfer (DSSAT) simulation model is used to estimate crop yield for the period from 1981 to 2025. The datasetsfrom India MeteorologicalDepartment (IMD) from 1981 to 2016 and National Oceanic and AtmosphericAdministration(NOAA)RCP4.5climatic variables from 2017 to 2025 were used in the current study. Wecanprovideadecisionsystemthatisableto learn from the input variable and predict the plant growth and development in real time. Our result indicates that there is (R2 = 0.67) and RMSE of 281 kg/ha from data predicted from random forest andCrop Yield Estimation. The studymakesuseof MODIS data from Earth Observing System satellites namelyTerra(EOS AM-1)& Aqua(EOSPM-1)from NASA. Thestudyvalidatesthepredictedyieldby comparing their values with NDVI values. High NDVI valuesmeansmorevegetationandmoreyield.Hence, itcanbesaidthatdecisiontreesandrandomforests can be used in forecasting the crop yield. FarmMonitoringusingcloudnetworksystemisused to generate decisions on irrigation using real time data. Farm Monitoring using cloud network proposed will help farmers to plan an irrigation schedule in his farm according to the collected data. Proper scheduling of irrigation and fertilization is very important for proper developmentofcropsaswellasfarmers. This [2]system will collect the data as per climatic conditionfromfieldsalongwithweatherdatafrom web repositories and this data can be used to take several

effective decisions for increasing the production of crops. Farm Monitoring using Cloud Network system used cloud and big data functionalities to store and retrieve data from anywhere, with the help of repositories and datasets. Cloud

And Big Data concepts are used separately to predicttheproductionofyields. Thissystemcollects data through sensors and stores this data in cloud databasethroughinternet. Thenusingthebigdata predictiveanalysistechniquestoanalyzeandpredict the crops production.

The result of the experiment is evaluated by the average error rate of 10-fold cross validation of each data set for 10 runs. 5-fold cross validation is a process which divides the data set into 4 blocks. 4 blocks are merged for training data and the rest block for the testingdata. Tableshowstheaccuracyofclassifying abalone data set, horse-colic data set and soybean data set. Wetesteverydatasetforfourtimesandthen calculate the average accuracy. Every time we randomly choose the training data and testing data. Using the training data to generate the decision tree classifier and then classify the testing data.

Accuracy	Jowar	Bajra	Soybean
Test 1	25.17	55.56	41.28
test 2	30.20	67.65	34.49
test 3	29.43	70.32	11.13
Test 4	27.59	69.26	14.25
Avg.Accuracy	28.10	65.70	25.29

Thissystemisproposedtofulfillthe gapbetween farmers and technology in agriculture. It helps farmers to understand the requirements to grow the crops in critical seasons and avoid financial losses caused to them. It will be helpful to the farmers of different geographicalregions.The[3] proposedsystemcontains facilities such as collecting real time data that will help farmerstomakeprecisedecisions. Theproposed system is cost effective as the entire process is softwaremanagedandsensormonitored. Thissystem will help in improving better yieldWeather data from the meteorological department can be used along with the sensed data to predict more information about the future which will help farmer planaccordinglyandimprovehislivelihood.

The proposed system will store all the information in a centralized [4]cloud, which will be available to all usersatalltime.Managementofalldatarelatedto land, location, area; soil and land characteristics through centralized decision support systems. Management of all data related to land, location, area; soil and land characteristics through centralized decision support systems.

The proposed system will manage all data related to land, location, soil through centralized decision support systems. Proposed system will eliminate farmer's limitations about technical knowledge and resources. It will provide farmer's to use agriculture resources efficiently.

Agriculture plays a crucial role in India's economy and around 70% of people earn their income through it and alsoprovideslargeemploymentopportunities. The technological advancement has led to remarkable achievementsindeveloping Agriculturalbased software applications to get faster information. But, many farmers are still applying the traditional methods of farming and hence the result of productivity becomesverylow. Agriculturepredictionprocessfor organic and inorganic farms is an open issue and it depends upon weather, soil fertility, water, seasons and commodity prices, etc. Soil fertility factor is paramount important to maintain the crop growth and increase theproduction.Thesoilfertilitylevelshelp the farmers to identify the deficiencies in the soil, namely nutrient content, soil type, pH value, EC (Electrical Conductivity) value and soil texture and to choosetherightcropstoincreasetheproduction.In this work, as a novelty, the soil fertility level is predictedbyanalyzingthe VirudhunagarDistrictSoil information and recommendations are offered for crop selectionandsowingbyusingC5.0: Advanced Decision Tree(ADT)classifieralgorithm.Usingthis technique,anAndroidbasedmobilephoneapplications named as Design of Smart Information System (DSIS) applicationhasbeendeveloped. Theproposed application activates the Global Positioning System (GPS) to identify the user location.The performance of proposedmodelisanalysedanditiscomparedwith the existing classification model for agricultural data. Agriculture is the backbone of India's economy. It also provides large employment opportunities to the people. Nowadays, the cropcultivation becomes very low since many farmers are still following the traditional farming methods. So the farmers, governments, agricultural scientists and the researchers are exploring new methods to get high yield from the farmland. Government of India is concentrating on the agriculture industry by handling several activities to increasethecropproductivity(VincyaandValarmathi, In

India, the primary nutrients (Nitrogen, Phosphorous, and Potassium)areverydeficient intheagriculturalsoil. So, it is mandatory to take necessary action for overcomingthedeficiency byapplyingthecorrect requirement of fertilizers for the crop. Soil health is the combination of physical, chemical and biological activitiesofsoil. Theessentialgoalofanytechnology with respect to agriculture is to make the better crop production with many

Immediate and sustainable benefits (AthmajaandHanumanthappa,2016). Farmers are not only harvesting the crops and vegetables,buttheyalsoharvestthemassiveamount ofdata. Therefore,thereisaneedofbigdataanalytics to analyze the same. It works with large datasets and problematical processing is done by database management applications or traditional data processing tools. MapReduce is used to access the big data sets with a parallel distributed algorithm and it is implemented with a programming model. It performs two operations:oneisMap()andanotherisReduce(). Map()takescareaboutfilteringandsortingthedata and Reduce () concentrates on summarizing the data (BhargaviandJyothi,2011).Dataminingdiscovers theunknowninformationfromthedataset. Thesoil nutrient information, soil composition and soil characteristics are required for determining the soil fertilitylevel. Also,themeasureofsoilfertility indicatesthedeficienciesthatneedtotaketreatment are based on the soil test.

II. RELATED WORK

Thesimulationmodelsplayaverysignificant rolein the development of the agro-ecological and socio-economic conditions. Several applications are in use, based on data mining algorithms in the agriculture field.Decision Support Systems (DSS) are used to generate data for agriculture management systems viz., the pest management, farm management and crop managementsystem. Theperformanceofthese systemsislow.Hence,theutilizationofIoT based advanced techniques could improve the process of farmer decisions about crop fertilization.

Many authors have done their research using data miningtechniqueswithsmallinstances. Thedecision support system model, related to soil water balance, consider several factors as soil type, weather, canal network,andcroptype(Oadetal.,2009).Gholapetal. (2012) used Decision Tree algorithm for predicting soil fertility. Author collectedthedatasetfromthe privatesoiltestinglaboratory,Puneandusedselection andboostingtechniquestotunetheperformanceofthe J48 algorithm. Ad boost is one of the boosting techniquesusedinthisresearch

Theypredictedthe soilfertilitylevelandclassifiedasverylow,veryhigh, low, high, moderate, andmoderatelyhigh. Further, they improvedtheaccuracylevel as 96.73% after applyingselectionandboostingalgorithm.Preethaiet al. proposed new approach for soil classificationand predicted various factors with the help of decision support system.

Theauthormonitoredtheagriculturalactivitieswith thehelpofthisnewdecisionsupportsystem. Their study isinvolved with the CoimbatoreDistrict dataset and by using enhanced C4.5 algorithm, fast decision support process is achieved. In addition, they have conveniently learned important information on pest and disease control through mobile phone. For rural formers, mobile multimedia agricultural advisory system based call centre provides the required information

III. LITERATURE REVIEW

Farm Monitoring using Cloud Network system used cloud and big data functionalities to store and retrieve data from anywhere, with the help of repositories and datasets.CloudandBigData conceptsare used separatelytopredicttheproductionofyields. This systemcollectsdatathroughsensorsandstoresthis dataincloud[5] databasethroughinternet. Thenusing the big data predictive analysis techniques to analyze and predict the crops production. Cloud database is used as it is an efficient and well defined application. Cloud based system integrated aims at follows: Cloud database is used to store and share the crop information's. Itgives smartness with flexibility, predictability, scalability and optimization. It provides information for farmers in an economical and reasonable cost Withregardstoarablecultivating,IoT frameworkscan possibly give new bits of knowledge and give a state- of-the-art situational mindfulness with an a lot more significant level of spatio-transient granularity of checking[6]. Theybolsterthecomprehensionof elements impacting crop development and yields, which is exceptionally critical for a feasible agribusiness. Throughsite-explicitadministration,IoT frameworks help to fundamentally spare homestead assets and, in this manner, increment ranch yield. Moreover, IoT-based harvest observing additionally improves yield displaying and the nature of yield expectations. By and large, the rising computerized upheaval,especiallyIoT combinationintopresentday

agriculture,isakeyempoweringagentthatpermitstorobot ize numerous procedures and bolster them with important extra data.

SincethepremiseoftheIoT anchorfromyieldsto ranchers is sensor-based information gathering, sensor gadgets and their in-situ sending is major for the achievement of Smart Farm-ing. Such gadgets run from little, ease, and asset obliged sensors to complex high-precision sensor stages that could be over the top expensive. For a huge scale crop moni-toring, for the most part numerous sensors are required. Henceforth, from a financial viewpoint, the cost of individual sensors is vital for the arrivalon venture (RoI). In Wireless Sensor Net-works (WSNs), such modest gadgets are ordinarily utilized for en-vironmental detectingofgenuinelyquantifiableparameters. These are remotely interconnected and intended for enormousscaleandlonghaularrangements. Asthey are dominatingly battery-driven, they are in like manner profoundly asset compelled. In this way, singular sensors have constrained detecting exactness that is remunerated by the enormous number of working together gadgets consistently gathering ecologicaldata. Also,ground-basedWSNscanbe supplemented by remote detecting that depends on airbornesymbolism. Thisistypicallyobtainedby satellites or as of late by rambles and, among others, is utilized to determine data about harvest development. This is especially advantageous to estimateWSN-situated in-situ data to significantly bigger regions.

IV. EXPECTED OUTCOME

Saves time by allowing you to always keep track of critical operations via a smartphone or tablet no matter where you are or the time of day. Increases efficiency by reducing the risk of unwanted downtime and interruptions. Increases security since the system alerts users of unexpected events and unwelcome visitors. Reduces the risk of environmental effects because you canimmediatelyseewhensomethinggoeswrong. A robust, easy-to-use system tailored to the specific needs and prerequisites of farms.

Easy to install and easy to expand and customise to the special conditions at your farm. Can also be supplemented with other types of monitoring, such as mobile cameras and special detectors, e.g. at the diesel tank, along fences and to monitor temperatures in differentlocations. Thisenablesyoutogatherall critical information for your operations in a single system. You can monitor everything from a smart, user-friendly app on your smartphone or tablet.An affordable, flexible monitoring system that can be adapted to any type of farm.

IoT basedfarmingsystemcanproveveryhelpfulto farmers since over as well as irrigation is not good for farming.Thresholdvaluesforclimaticconditionslikehum idity, temperature, moisture can be fixed based on the environmental conditions of that particular region. This system will generate irrigation schedule based on the real time sensed data from field and data from the weatherrepository. Thissystemcanrecommend farmers weather or not there is a need for irrigation.

IoT servesasapowerful,reliable,andcosteffective technology tto implement idea of “” that aims to empowerment of farmers with advanced connectivity through web services, measurement of environment factors like soil moisture, temperature, humidity and implementingcloudcomputingalongwithrealtimemonit oring

[6]RaghuviraPratapA,KSuvarnaVani,AnEfficient Density based Improved K- Medoids Clustering algorithm,*InternationalJournalof Advanced ComputerScienceand Applications(IJACSA),Vol.2, No. 6, 2011.*

[7]NikitaP.Katariya1,M.S.Chaudhari, Text PreprocessingFor TextMiningUsingSide Information,*InternationalJournal of Computer ScienceandMobile Applications(IJCSMA),*Vol.3, January- 2015, pp. 01-05.

[8] K.Sruthi, B.Venkateshwar Reddy, Document Clusteringon VariousSimilarityMeasures, *InternationalJournalof AdvancedResearchin Computer Science and Software Engineering (IJARCSSE),* Vol. 3,August 2013.

[9] “Microcontroller Based Drip Irrigation System Using Smart Sensor” by Nilesh, R. Patel, Rahul B. Lanjewar in the year 2013.

[10] [WaleedAl Shehri,’CLOUD DATABASE - DATABASEASASERVICE’, International Journal of Database Management Systems (IJDMS)Vol.5, No.2, April 2013]

[11] [Harrison John Bhatti ,Babak Bashari Rad , Databases in Cloud Computing:ALiterature Review, Article in International Journal of Information Technology and Computer Science ·April 2017]

[12] [Shweta Dinesh Bijwe, Prof. P. L. Ramteke Database in Cloud Computing - Database-as-a Service (DBaas) with its Challenges,Shweta Dinesh Bijwe et al, International Journal of Computer Science and Mobile Computing,Vol.4 Issue.2, February-2015, pg.73-79]

[13] https://www.codeigniter.com/user_guide/general/requirements.html

[14] https://codeigniter4.github.io/userguide/ changelogs/v4.0.0-alpha.4.html

[15] https://www.techopedia.com/definition/27209/base64

[16] https://www.tutorialspoint.com/mysql/mysql- introduction.html

Detection ofPhishingwebsites usingmachine learning

AnkitYadav, AshwinYadav, YashArohan
Department of Comuter Engineering
Thakur College of Engineering and Technology
Mumbai, India
ankit.yadav2597@gmail.com, ashwindyadav198@gmail.com, yash.arohan@gmail.com

Abstract—**Web serviceis oneof thekey communications softwareservicesforthe Internet.Webphishingisoneofmany securitythreatsto web servicesonthe Internet.Phishingis amongoneofthemost vulnerablesecurity threatwhichposesa problemto the internetusers.Phishingattackcan bebroadly classifiedassocialengineeringattackwithamaliciousintent, theattackeradvertentlycreatesafakewebsitewhichtempts the users click it forging a straight path for toward the sensitiveinformationofusers.Thereare amanywaysto distinguishphishingwebsitesor phishingURLsfromthe authenticones.Phishingwebsitescontainmultiplecuesboth in their content part a s well as i n browser-based security indicatorswhichisnearlyimpossibleevenforan expert perpetratorto circumvent.Therearemanysolutionsproposed byseveralresearchersinwhichnosingleapproachevolvesas aneffectivemethodtodetectphishingattacks.Cyber security personnelare lookingfortrustworthyandsteadydetection techniquesand toolstodetectphishing.Weaimtotacklethis problemwith growingmachinelearningtechnologyby extractingand analyzingvariouskey featuresoflegitimateand phishingURLs.DecisionTree,amachinealgorithm,notable foritsclassificationquirk,areusedto detectphishingwebsites. Ourmainaimis to detectphishingURLsand furthernarrow down tobestusecasemachinelearningalgorithmbasedon accuracy, false positive and false negative rate of each algorithm.**

Keywords—Decisiontree,URLs,machinelearning,phishing websites,phishingfeatures.

I. INTRODUCTION

Phishingis acybercrimeemployingbothsocialengineering and technicalsubterfuge to trickconsumers'revealingtheirpersonalidentity,data,a ndfinancialcredentials.It is known to become numberoneattackvectorsurpassingother web-application-attacks.Socialengineering schemes use mediumsuch ase-mails,links andattachments; purportingtobefrom legitimatebusinessesand agencies, designeddeliberately to leadconsumers tocounterfeit websites thattrickrecipients into divulgingfinancialdata such asusernamesand passwords. Bamboozling schemes suchasthisis merelyafirststepinattacker'scourseof action,it couldfurtherinstallmaliciousssoftwareonto computers,togainfullaccessdirectly,often usingsystems to interceptconsumers'onlineaccountuser namesand passwords or act as a chain in command of botnets, makingitmostcommonandreadily availablecybercrime vector. Withadvancementsintechnology andthe perpetrator naivetéhoodwinkvictims,an effectivesolutionto counter-attackthiswidely common attack came intolight. Machine Learning, hasa broad spectrum and can be applied everywhere if there is achanceofimprovingthesolution based on itsprevious solution.Theprimaryobjectiveis todifferentiatephishing websitefromauthenticones andsummarizewithwhat algorithmworksbestwiththetestcases. Statisticshaveshownthatthenumberofphishingattacks Keepsincreasing,whichpresentsasecurity risktotheuser informationaccordingto theAnti-PhishingWorkingGroup (APWG)[1]andrecordedphishingattacksbyKaspersky Lab[2],whichstatedthatithasincreasedby47.48%from allofthephishingattacksthathavebeendetectedduring 2016

II. BACKGROUND

Phishingisoneof themajorproblemsoftheinformation security.Itcanoccurintwoways,eitherby receiving suspiciousemailsthatleadtothefraudulentsiteorby users accessinglinksthatgodirectly toaphishingwebsite. However,thetwomethodsarecommoninonething,which isthe attackertargetshumanvulnerabilitiesratherthan softwarevulnerabilities.Phishing canbedescribedas fraudstersthattry tomanipulatetheuserintogivingthem their personal informationsuchasusername,password,anda creditcardnumber.Thesescamsareleadingtoeconomic andfinancialcrisesforusers[4].

Recently,there havebeenseveralstudiestryingtosolve thephishing problem. Theycan becategorizedintofourcategories:blacklist, heuristic,contentanalysis,and machine learning techniques.TheblacklistcomparestheURLwithan existingdatabase thatcontains alistof phishingwebsite URLs.Becauseoftherapidincreaseofphishingwebsites, theblacklist approach hasbecome inefficient in deciding whethereachURLisaphishingwebsiteornot,andthiskind ofdelaycanleadtozero-dayattacksfromnewphishingsites [4]..This templatehasbeentailoredforoutputontheA4 papersize.Ifyou areusingUSletter-sized paper,please closethisfileanddownloadtheMicrosoftWord, Letterfile.

This model present itself as a stopgap until more sophisticatedalternativescomewithbetterreliability,rather itsets afootinthefieldof cyber-security toobstructatleast thenaïveattacksdonebynewcomerstogain confidence which mayultimately pose a greater threat. The

trainedalgorithm boasttodistinguisheventhe newly formed domainURLandcanperceivethemasthreatif theypose themselvesason.Itwillplay animportantroletobothold andyounggenerationunawareofacyber-worldofwhich they areapart.Uneducatedpeopleorpeopleunawarewere onlyleftwithclueswhenposedinsuchsituationandhadno

withtheirinexperience.Despiteitsdistinguishing attributes evengiantcompaniesfallpreytosuchattackvectorforgin g astraightpathtosurplusdata.

III.RELATEDWORK

A. Content BasedApproach

Zhangetal.[3]presentedthe designandevaluationof CANTINA,anovel,content-basedapproachto detecting phishingwebsites,usingthewell-knownTF-IDFalgorithm. Itanalysesthe text-basedcontentofapagebyitself.They experimentedwithsomesimpleheuristicsthatcanbeappli ed to reducefalse positives. As a result, a pure TF-IDF approachcancatchabout97%ofphishingsiteswithabout 6%falsepositives,andwithheuristicsitcatchabout90%of phishingsiteswithonly 1%falsepositives.Rao andAli[5] implemented a desktop application to detect phishing websitesusing anovelheuristicbasedonURLsandwebsite content.

B. URLAPPROACH

Pradeepthi.K V andKannan.A[11]hadproposedto detectphishingsitesisby derivingandconcludedthattree- based classifier are best and gives good accuracy for phishing URL detection. Also they have usedvariousfeaturessuchaslexicalfeatures,URLbased feature,network basedfeatures anddomainbasedfeature.Itgiveslow predictionaccuracyforadatasetascomparedto other machine learning algorithms. Jeevaand Rajsingh[8] presented asystemforpredictionphishingURLsby generating rulesofassociationrule mining.Theyusedthe apriorialgorithmto pickknowninformationfromfrequent item setproperties that were extracted from the dataset. JeevaandRajsingh[8] alsousedanotheralgorithmthat performsonhiddendata toobtaintheaccuracyofassociation rules, which is apredictive apriorithat engages the confidenceandthesupporttechniquesthataremeasuredi n itsaccuracy,unlikeapriori,whichonlymarkrulesthathav e theconfidencetechnique.

C. MACHINELEARNING

PriyankaSingh,YogendraP.S. Maravi,SanjeevSharma [9]proposedanovelapproachby combining twoormore algorithms.Here,theauthors haveimplementedvarious algorithmsalongwithSVMforgettinggooddetectionrate and classificationpurpose. Themain disadvantageof theSVMalgorithmisthatithasseveralkeyparametersthat idea whattheygot themselvesinvolvedinto. Hence theprojectcontributestothegrowingworldanditsthreatwit

Machinelearning model.LikePhishtank [6],opensource community bolsteringandstrengtheninganti-phishing,the projectisalsomeantforthegullible societyandcopeup

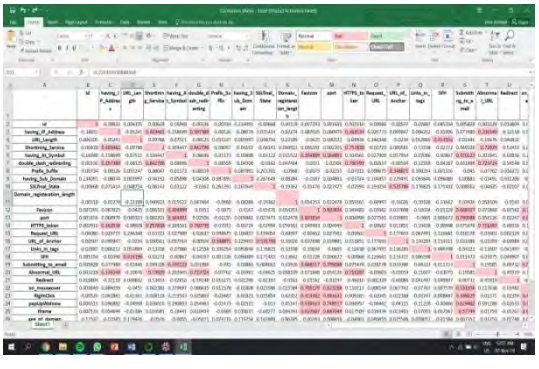
*A. Dataset*IV.EXPERIMENT

needtobesetcorrectlytoachieve thebest classificationresultsforanygivenproblem.Parametersthatmayresu ltin an excellent classification accuracy for problem A, may resultinapoorclassificationaccuracy forproblemB.Fresh-Phish:Aframework forauto-detectionof PhishingWebsites [10].Inthis,they haveused reducedfeaturessetandusing pythonforbuildingquery. Theybuilda largelabelleddataset and analysesseveralmachine-learningclassifiersagainstthis dataset. Analysis of thisgives verygood accuracyusing machine-learningclassifiers.These analyseshowlongtimeit takestotrainthe model.This modelhasanaccuracyof90% whichislowwhencomparedtoothermodels

Wecollect16000of phishing and legitimateURLs.The phishingwebsitesconsistof12000phishingURLsthat has beencollectedfromPhishTank[6].Inthe otherhand,the legitimate websitesconsistof 4000 legitimate URLsthathavebeencollectedby adailyusefrom10chosenusers. However,thefinaldatasetafterhandling missingdataand removingtheduplicateissizeof6116.

B. FEATURESEXTRACTION

Thephishingwebsiteshavecertain characteristicsand patternsthatcanbeconsideredasfeatures.Inthissubsectio n, wecoverallphishingwebsitefeaturesthat havebeenusedin thepreviousresearchesaspossible.Furthermore,whilewe arestudyingthe phishingcharacteristicsandpatternswe noticesome new characteristicsthatcanbe consideredas features.The totalnumberofphishingfeaturesis36where3 ofthemarenewfeatures. Wecategorizethemintothree maincategoriesasshownbelow withfeaturesin table I:• Featurescanbeextractedfrom URL.• Featurescanbe extractedfrompagecontent.•Featurescanbeextractedfro m page rank.Weusethenumberofinputemailandnumberof inputpasswordasthenewfeatures for phishing website, sincethe targetofthe phishingwebsiteis tostealsensitive informationsuchas emailandpassword.Weconsiderthe numberofinputthathavethe typeemailorpasswordas featureforphishingwebsite.Anothernewfeature is the numberofbutton,whilewearestudyingphishingfeatures wenoticedthatalarge numberofphishingwebsitedoesn't usethesubmit buttoninsteadtheyusearegularbutton,sowe considerit asfeatureforphishingwebsite.



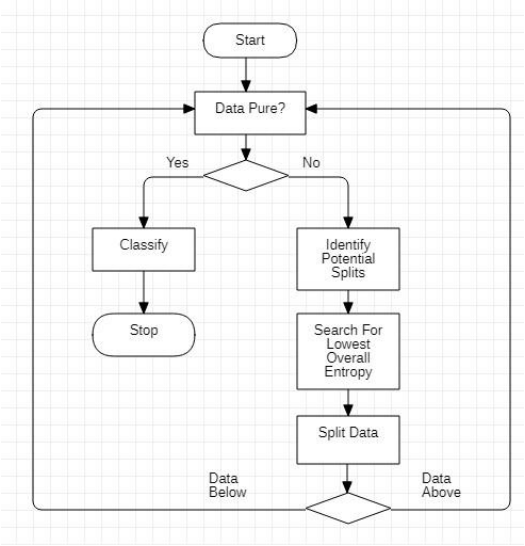
TABLEI. FEATURESTHATCANBEEXTRACTEDFROMTHEURL

LengthofURL	LengthofhostnameofURL
Lengthofthe pathofURL	Numberofdot (.)in the path
Numberofdot (.)in hostname	Numberofslashes(/)in URL
Numberofhyphen(-)in hostname	Numberofspecialcharacters(:; % &?+)
Numberofat(@) in theURL	Numberofdigitin hostname
Numberofunderscore()in hostname	Numberofunderscore()in path
Numberofcertainkeywordin URL	Numberofhexadecimalwith%
Transportlayersecurity	IPaddress
Presenceof www	Portredirect
UnicodeinURL	Hexadecimalcharacters

C. DecisionTree

In this project we will be using decision tree as our supervisedmachinelearningalgorithmbecauseofitsfast and reliable resultsandeasy loadonhardwareto makeit robust at the same time. It drawstaxonomyof allthefeatures by reading and calculating their entropy o

Heuristics,likelinessorplausibility,ofallandtheir recurring featurestoformatreelikestructure.



V.RESULT ANDCONCLUSION

Wehavestudied all30featuresandtheirco-relation.Butto underweightfew important,the featuresetwasreducedwith thehelpofPCA.The projectspreprocessesthedatasetwith PCAalgorithmtoenhance thecopiousamountthrough dimensionreductionto makeitsoundandcoherent.Italso removesnoiseandinappropriatedatavalueswhichwouldh e ahindrancethelearning process

Phishing attacksthroughURLsisoneofthemajorissues facedby theInternetCommunitybecauseoftheonline transactionsperformedonaday-to-day basis.Thephishing attackscauseseverelossto companies,customers andweb users.SocialnetworkingsitessuchasFacebook,Twitteran d LinkedInhavebeenthevictimsof phishing.However,there areanti-phishingtools available whichcanhelpusersto detectphishingattacksandpreventthem.TheDecisionTr ee based Phishing Detection System detects the malicious URLs which will help the users to be aware of such maliciousandsuspiciousURLs.Aftertheuserentersa URL,thesystemprovides16featuresforthe usertoselect andapply oninputURLto determinewhetheraURLis maliciousorlegitimate.

VI ACKNOWLEDGMENT

WesincerelythankourguideMs.PrachiJanraoforher valuableadviceguidanceandconstantsupportandalsofor thestick toourbacks.

References

[1] AOKasperskylab.(2017).TheDangersofPhishing:Help employees avoidthelureofcybercrime. [Online]Available: https://go.kaspersky.com/Dangers-Phishing-Landing-Page-Soc.html [Oct30,2017].

[2] ”Financialthreatsin2016:EverySecondPhishingAttackAimsto StealYourMoney” Internet: https://www.kaspersky.com/about/pressreleases/2017 financial-threats-in-2016.Feb22,2017[Oct30,2017].

[3] M.Blasi,”Techniquesfor detectingzeroday phishingwebsites.” M.A.thesis,IowaStateUniversity, USA,2009.R. Nicole, “Titleof paperwithonlyfirstwordcapitalized,”J.NameStand.Abbrev.,in press.

[4] ”PhishTank—Jointhefightagainstphishing.”[Online].creasing ProblemofElectronicIdentityTheft.Wiley,2006,pp.2–3.

[5] M.Blasi,”Techniquesfordetectingzerodayphishingwebsites.” M.A.thesis,IowaStateUniversity, USA,2009.

[6] L. A.T.Nguyen,B.L.To,H. K.Nguyen,andM.H.Nguyen, ”DetectingphishingwebsitesAheuristicURL-basedapproach,”in 2013 International Conference onAdvanced Technologies for Communications(ATC2013),2013,pp.597-602.

[7] Z.Zhang,Q.He,andB.Wang,”ANovelMulti-LayerHeuristic ModelforAnti-Phishing,”NewYork,NY,USA, 2017,p.21:1-21:6.

[8] N. Sanglerdsinlapachaiand A. Rungsawang, ”Web Phishing Detection UsingClassifierEnsemble,”NewYork,NY,USA,2010, pp.210-215.

[9] Singh,P.,Maravi,Y.P.S.,&Sharma,S.(2015). *Phishingwebsites detectionthroughsupervisedlearningnetworks. 2015International ConferenceonComputingandCommunicationsTechnologies*

[10] Shirazi, H., Haefner, K., &Ray, I. (2017). *Fresh-Phish: A Framework forAuto-Detectionof Phishing Websites. 2017 IEEE International Conference on Information Reuse and Integration (IRI)*.doi:10.1109/iri.2017.40

[11] Pradeepthi,K.V.,&Kannan,A.(2014).*Performancestudyof classificationtechniquesforphishingURLdetection. 2014Sixth International Conference on Advanced Computing (ICoAC)*.doi:10.1109/icoac.2014.7229761

ImplementingSystemforLocationBasedMiningforWaterTreatmentinFarm

Ms. Prachi Janrao, Abhay Pandey, Akash Singh, Vikas Yadav
Department of Computer Department
Thakur college of Engineering and Technology
Mumbai, India
prachi.janrao@thakureducation,org, abhayworks07@gmail.com, as1470601@gmail.com, vikas071998@gmail.com

Abstract— **Currently about 70-85% of total irrigation of agriculture inIndia isdone through borewell and tubewellsi.e.viathegroundwater. Groundwaterishard waterwhichhas ahigh concentrationofminerals,andofthesecalcium salts arethemosttroublesome.Oneofthemain element which isfound inabundant and which ismainly responsible for causing hardness is calcium carbonate.**

Thesemetalsand hardness causingagents are harmfulofhumanhealthas wellas foragriculture. Althoughtheeffectisnotthatprominent, soitis ignoredmostof thetime. Butona longrun ithasa veryhugeimpactonthefieldofthecropandthe overall agriculture.

Thisproblem ofhardwatercanberesolvedby reducingthe concentrationofthesemetalsand salts by electro-magneticwater softener.Thiswater softenerisnotaburdentothepocketas itis affordable for the farmers. It operates at a frequencyof15kHzandpoweredbya 5-9voltage powersupply.This refinedwaterwillincreasethe fieldtoagreatextend.

Keywords—Calcium Carbonate,ReverseOsmosis.

I. INTRODUCTION

Withthecurrenteconomicalconditionand unstable market,thecategoryofpeoplethatsufferthemost arefarmers.Theyworkhardday in-dayouttoearn fortheirlivelihoodandintheendtheyarejustableto managefoodfortheirfamilyandsavesomepenny forfuture.Variousthingscanbedonetoimprove thisconditionoffarmersandoneof themistohelp andimprovethefieldoftheircultivation withno muchburnedontheirpockets.

Whenwe lookatthe parametersandthe possible reasonsfortheyield tobenot that significant,soone ofthefactor onwhichitcandependon agreat extendistheconditionandquality ofwaterwhichis being used for agriculture. In India mainly, agricultureisdonewiththe helpofgroundwaterfor mostof theperiodinayearandinthemonthsof monsoonthe farmersdependonrainasasourceof water fortheircrops.

Theproblem withgroundwater isthatitcontainslots of metals,saltsandhardnesscausingagents.This issueis ignoredtoagreatestextendwhenit comesto farming,butthesesaltsand hardnesscausingagents arenot goodforagricultureandit hinders thegrowth ofcroptoagreatextend.If thehardnessisremoved sotheoutcomeoftheagriculturewillbe more than expectedandconditionoffarmerswillbeimprove a andextend.

Ifthe hardness isremovedtheoutcomewillchange drastically,sotheproblemarisesishow toremove hardnessand saltofthewater atcheapandaffordable way.Therearevariousfiltersthatuse chemicaland electro-chemicalmethodstoremovehardnessof the waterduringitsbeingusedandthey arevery expensive. Thereisa oldtheorywhich saysthat whenhardwaterispassed throughastrongmagnetic fieldtheCaCO3elementscombinetogethertoform asinglemolecule.

Withthehelpof abovetheoryweuseelectro-magneticmethodto removehardnessandmakeit evenmore-betterforagriculture.This processesis goingtobeachievedby creatinganstrong electromagneticfieldandthewaterwillbe passed throughthisfieldwhichwillhelptoreduce the hardness andwaterwillbebetterforuse.Notonly combiningthehardnesscausing agentswillsolvethe problem but even filteringoutheseimpuritiesis equallyimportant.

Thisprojectusestheconceptofextractingdataabout theconditionofwaterof particularlocationand dependingonthatvaluetheamountofhardnessfrom thewaterwillberemoved.Thisprocesswillnotcreate anunnecessary stressonthecircuitand itcanworkin proper condition ofalongtime.Thiscircuitwillbe connectedwithafiltertoremovetheimpuritiesfrom thewatermakingitsconditionbetterthanbefore.

II. LITERATUREREVIEW

[1] EvaluatingGroundWaterQuality andSuitability forIrrigation.ApplWaterScience.By DianaAnoubam Sharma,Madhuri S. Rishi, TirumaleshKeesari.Explains how thelevelof impuritiesreducestheyieldofthefarm.Butthe gapishowtoremovetheimpuritiesofthewater.

[2]AnalysisofGroundWaterQuality Parameters. ByDevendraDohare,ShriramDeshpandeandAtul Kotiyan.Explains thevariousimpuritiespresentin groundwatercausinghardness.Butthegapis how toremovehardnessfromwater.

[3] Design and Construction of an Electronic WaterSoftener.By AngelicaA.GaberialandB.S. Jatau.Explainshowtoreducehardnessofwaterby electricalmethod.Butthegap isthatitdoesnot removeotherimpuritiesinvolvingsalts

[4] Colpitts oscillator: Design and Performance Optimization.By AbkitRana,PoojaGaikwad. Explains how to create and the principles of colpittsoscillator.Butitdoesnotexplainhow can this circuit be implemented to remove hardness fromgroundwater.

[5]ProcessandApplicationof ReverseOsmosis. Explainshowtousetheprocessofreverseosmosis forfiltrationofwatercontaining impurities.Butthe gapisthat,itdoesnotspecifytheuseoffiltration inagriculture.

III. EASEOFUSE

Thisprojectwillhelpfarmersto reducetheirefforts, expenseby improvingtheiroveralloutcome.Thecrops willgetabetteryieldand thefarmerswillearnbetter returns.Itaimsto removehardnessfromthewaterand increaseitsefficiencyandusability.

IV. HOWITWORKS

It works on the principle that when hard water is passedthroughastrong magneticfieldthehardness causingelementscombine togetherforforma comparativelylargermolecule.Thiscanbefilteredout later using ROfilter. For theprocess of creating a strong magnetic field the project will use colpitts crystaloscillator.Theoscillator willbedesignedin suchaway thatitoperatesonafrequencyof25MHzThefrequency willbecontrolledusingavoltage regulatorwiththehelpof Adriano.TheAdrianowill determinethevoltageforthecircuitdependingupon the locationandapplydataminingforthesame.

Figure1:AdrianoUNOBoard

Output pinvoltagecanbefluctuatedusing analogwrite function of AdrianoUNO. This helps in reducingunnecessary hardtocircuitcomponentsand savespowerconsumption

A. Abbreviations andAcronyms

CaCO3-*CalciumCarbonate*

RO-*ReverseOsmosis*

LC*-InductorCapacitor*

B. Algorithm/Circuit

Theprojectwillusecolitiscrystaloscillatorwhichuses aLCtankcircuitto produceafeedback.Thecapacitor dividergeneratesapositivefeedbackto generatean oscillationwave.This tankcircuitwillgeneratean oscillator frequencyof26kHzwhichisrequiredfor combiningCaCO3 moleculestogether.Crystalis connectedinparallelwith resistors andcapacitorbiased junction.Thepowersupplywillbegivenby Adriano dependinguponthe locationofuse.Thecircuitdiagram isshownbelowwhichshows alltheconnection.

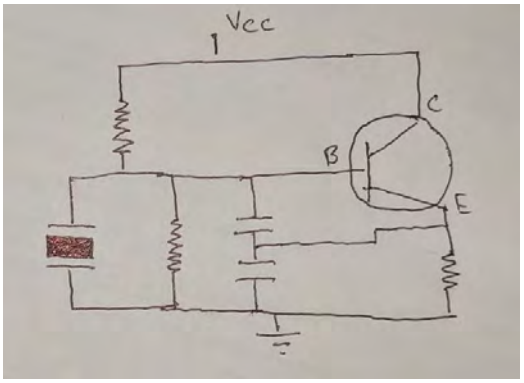


Figure2:ColpittsOscillatorCircuit

C. Applications

Thisprojectwillbeusedbyfarmersto reducethe hardnesscausingagentsfromthegroundwater.This will improve cultivation of farm as hardness is the waterisoneof themainhindranceforcultivation.

Farmerswillgetan overallpositivereturn fromthis project.

D. Advantages

Themainadvantageof thisprojectisthatthehardness causingagentsfromthegroundwateris removedand waterisfreefromimpurities.The projectwill helpthe farmersto getabetteryieldandbetterreturnsfortheir cultivation.

E. Methodology

Inourprojectwearegoingto useagileprocessmodel becauseitisincrementalaswellasiterativeinnature andit canadopt changesquickly.

Agile methodology isa practice that promotes continuous iterationof development and testingthroughoutthesoftwaredevelopmentlifecycle of theproject. AgileMethodsbreaktheproductinto smallincrementalbuilds.Thesebuildsare providedin iterations.Eachiterationtypicallylastsfromaboutone tothreeweeks.Everyiterationinvolvescrossfunctional teamsworkingsimultaneously onvariousareaslike– Planning,RequirementsAnalysis,Design, Coding, Unit Testing andAcceptanceTesting.

We aregoingto divideourprojectintodifferentphases andgoingtoworkonthephasessimultaneously so that the project can be done within stipulated time with properefficiency.Wearegoing to useAgile methodologybecauseitcanadoptchangesquicklyand if therewillbebuginthecodethenwecanfixthebug simultaneouslywhiletheprojectisgoingon.We can check our project and can make the decision accordingly.Whenwe aredone withparticularphase thenwecantestthesamephasesimultaneously.

F. Objectives

Todeterminetheagentsthatcause hardnessinwater.

Implementasystemtoremovethis hardness andincreasethefieldof agriculture.

Implementdatamining forremovalof hardnessinthewaterdependinguponthe location forwhichtheprojectisbeing used.

Theprojectwillrequireverylesspowerto operateanditwillnotbeaburdenforthe farmer

Theprojectwillnothaveanyadverse effecton farmoronenvironment.

Filteroutttheimpuritiesfromthe water andmake itinthebeststate inwhichit canbeuse

V. FUTURESCOPE

The project will use date set for all the location available from respective government websites. This datasetwill helptodeterminethe hardnesspresentin groundwaterat allthepossiblelocations.Thisproject willbeableto setthevoltagevalueforallthelocation withoutanyextrainformation.Thewayofreducingthe timecomplexity fromtraversingthousandsofrecordsin thedatabasewillbesolvedby findingmostappropriate algorithmdependingonthedataset.

VI. ACKNOLEDGEMENT

WesincerelythankourguideMs.PrachiJanraofor her guidanceandconstantsupportandalsofor the stickto ourbacks.We alsothankstheproject coordinatorsforarrangingthe necessaryfacilitiesto carry outtheprojectwork.

We thank theHOD,Dr.SheetalRathi,DeanAcademic, Dr.R.RSedamkar,thePrincipal,Dr.B.K.Mishraand thecollegemanagementfortheirsupport

References

[1] Devendra Dohare, Shriram Deshpande andAtulKotiyan(2014)Analysisof Ground WaterQuality Parameters:AReview.Research Journal ofEngineering Sciences

[2] Agelaga A. Gaberial and B. S. Jatau (2012)DesignandConstructionof anElectronic WaterSoftener.InternationalJournal of Engineering andScience..

[3] GarudR.M.,KoreS.V.,KulkarniG.S. (2011) Process andApplicationof Reverse Osmosis.UniversalJournalof Environment ResearchandTechnology.

[4] DianaAnoubamSharma,MadhuriS. Rishi, TirumaleshKeesari(2016) EvaluatingGroundWater QualityandSuitabilityforIrrigation.ApplWaterSci. [5] Luzia Grigorescu, Ioana Diaconescu (2016)

Theoretical and Experimental research on the current-modeRCoscillators.IEEE.

[6] Abkit Rana , Pooja Gaikwad (2014) Colpitts oscillator:Designand Performance Optimization. International Journal ofApplied

OptimizedSystem forMumbai Auto Rickshaw Drivers

Brij Parekh, Shashank Pandav, Ashish Yadav
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
parekhbrij16@gmail.com, shashank.pandav3@gmail.com, ashish1998.work@gmail.com

Abstract— **Thisapplicationis amobilityserviceprovider. An application that matches passengers withrickshaws, via websitesandmobileapps.Itisoftenhardto findarickshawin Mumbaiwhenneededthe most.Thisapplicationmatchesyou witha rickshawwillingtotakeyoutoyourdesireddestination. Toorderarickshaw,itisnecessarytoownasmartphoneand to register within the mobile application by entering your name,e-mailaddress,a cellphonenumber.Globalpositioning systeminthesmartphoneisusedtodeterminethelocationso thepassengerdoesnothaveto knowtheexactpickupaddress. The ride order appears on thenearest driver's smartphone applicationand he/shecanacceptorrejectthe ride.The applicationaimsto bridgethegap betweenrickshawsand passengers andprovidesthemwithaplatform tofindfares easily, without thefear ofconvoluted interfaces likeOlaor Uberwhichoften scarethedrivers.**

The app does not interfere withthe payment system of rickshawdrivers,making iteasyforanyrickshawdriverto adaptunlikeothermobilityserviceproviderswhich restrictthe driversto followthecompanyprotocols.Thispaperdescribes theapplication,itsdevelopment,andits technical implementation.

Keywords—Mumbai, Mobility Service Provider, Global PositioningSystem,Rickshaw,Smartphones.

I. INTRODUCTION

Itisverydifficulttofindanauto rideinMumbaiwhen neededthe most.Thisprojectfocusesonsolvingthis issue. Basically,itis anapplicationthatconnectspassengerswith Rickshawdrivers.

To order an auto rickshaw, it isnecessaryto own a smartphoneandtoregisterwithinthe mobileapplicationby enteringyourname,e-mailaddress,acellphone number. Globalpositioningsysteminthesmartphoneisused to determinethelocation sothepassenger doesnothaveto knowtheexactpickupaddress.Theride orderappearsonthe nearestdriver'ssmartphoneapplicationandhe/shecanacce pt orrejecttheride.

The app does not aim to interfere with the payment systemofrickshawdrivers,makingiteasyforany rickshaw driverto adaptunlikeothermobilityserviceproviderswhich restrictthedriverstofollow thecompanyprotocol.

Generally, wefocusonsavingpeople'svaluabletime, promotethe useofsustainabletransport,optimisetrafficand givingthe autorickshawdriversacomparativelymoreuser friendlyUI,freedomtouse

traditionalmethods andan opportunitytooptimisemonthlyprofit.

II. MOTIVATION

Withover110,000autorickshawdriverscatering toaround2 millionMumbaikarsondaily basis,wearestillfacingan immenseproblemoftransportinthecity.There hasbeenan exponentialincreaseinthe riseofgiantlikeOlaandUber especiallyinMumbaibutthisproblemisunsolved. This motivatesus tobringinasolutiontooptimizeetheuse of publictransportoralready existingfacilitywhichcould eventuallycontributeto thesolvationoftrafficproblemsin Mumbai.

III. PROBLEMDEFINITION

Giants likeOlaorUber recentlycameupwithOlaAuto andUberAutobutbecauseof theirbyzantinebusiness models,it becameworse.

The majorproblemsfacedbytheautorickshawdrivers includedhighcommissionrates,complexUI, usageofonline paymentmethods.It turnedoutthatthey made20%more profit goingthetraditionalway.

Bymakingthisapplication,weaim tosolvealltheabove problems andalsofocusonreducingthewastageoftimethat goesinvainfindingrides.

IV. OBJECTIVESOFTHEPROJECT

Ourmainobjectiveofbuildingthisapplicationistoact asbridgebetweenthe autorickshawdrivers inMumbaiand users tooptimisethe ridesatthe timeofrushhourandalso ensuring thateveryuserisableto gettherideontimeand everydriverisabletogetrideaccordingtohispreference.

Overall,objectivesofthisprojectareto:

- PromoteSustainableUrbanTransport
- Createamulti-lingualapplicationwithasimpleruser-interface
- Createanapplicationsupportingtraditionalpayment methods
- Optimizetraffic

V. SCOPEOFTHEPROJECT

Having110, 00Autorickshawdrivers and200,000 auto commentatorsoneverydaybasis.Withthesehugenumbers, itissignificantthatwehave agiganticmarkettotargetand thisapplicationhasacommendablescopeandweareveryOptimisticaftertheresponse thisideagot in themarket analysissurvey.

VI. APPLICATIONOFPROJECT

- These arethemajorapplicationsinclude:
- PrivateAutorickshawService
- SharedAutorickshawService
- 200,000Commentators: Offices,Schools,Colleges, Workers, Hawkers,etc.
- VirtualAutoRickshawService
- TrafficOptimization

VII. MARKETANALYSISSURVEY

AMarketAnalysisof about80 autorickshawdriverswas doneenquiringaboutthefollowingquestions:

- Areyouanownerorauto-rickshawisonRent?
- Areainwhichyou drivemostly
- Type ofdriver: Ola,Uber,Share,Private
- According aSurvey,most ofthepassengerstravel by Ola/Uberbecauseof therejectionsby auto-drivers.Whatexactlyistheproblemfacedby auto-drivers?
- Accordingtoyouwhatisthemajorproblemfaced byOla/UberAuto-rickshawdrivers.(Thisquestion is only applicable if Ola/Uber driver. If normal driver,put NA)
- Howmanyhoursadaydoyoudrive?
- Howmanyhoursyouwastefindingaride?
- Howmanyhoursyouwastefindingaride?
- Ifthere existsa solution toalltheproblems you discussed you pay on a monthly basis. May be aroundRs.500orhow muchyoucouldaffordtopay monthly?

Accordingtothe survey,56.2%ofauto-rickshawdrivers rentedtherickshawwhiletheother43.8%were theowners

Whilethehighestpercentvotedtheyfacedproblemswith paymentsystemsofOla/UmberAuto, alarge proportionvoted theydidn't understandtheconvoluted userinterfaces

VIII. PROPOSED WORK

The perceptionsofauto-rickshawsandtheirdriversamongst the middleclasses,media,consumerorganizations,and policymakersarelargely negative.Thevehiclesareseenas polluting,unsafe,and asignificantcauseofcongestion,and the drivers as greedy.Auto-rickshawsare perceivedasan intrusion into o r d e r e d u r b a n space, to b e t a m e d a n d controlledwithstrictpoliciesand punitivepenalties. Policymakingrelatedtoauto-rickshawsis dysfunctional–it doesnotadequatelyaccountfortheperspectiveofdrivers norforthefactthatthey provideа much-valued,cost-effective transportation service. While the concernsregardingovercharging,for example,onthepartofauto- rickshaw users, are understandable, policymaking has focusedlargelyon enforcingcompliancewithregulations, whichdo notreflectoperatingconditions.Auto-rickshaw numbersareseverely restrictedtocontrolpollutionand mitigatecongestion;private motorvehiclesare not,despite their rapidgrowth.

Awiderangeofpolicy responsesandtechnological advancementsare necessarytorespondtothecriticismsand negativeperceptions regardingauto-rickshawsandtheir impacts,toimproveservicequalityforauto-rickshaw users, andtoadvancethewelfareofauto-rickshawdrivers(note that theseobjectivesarenot antitheticaltoeachother,butare mutuallysupporting).Thisisparticularly necessaryifthe auto-rickshaw is to compete against new entrants in the publictransportmarketwhoutilizesmartphonetechnology tolinkpassengerswithindividualdrivers whodrivecarsor smallvans,whichare,unlike auto-rickshaws,notexposedto dustandweatherand areoftenair-conditioned.Thesenew competitors,suchasOlaandUber,are increasinglyavailable inurbanIndiaandoffermorecomfortableservice for above auto-rickshaw rates.Fixedpricesalsoremovehagglingand theperceptionofover-chargingandindividualgreedfrom thetransaction.The entranceofthesecompetitorsprovides auto-rickshawdriverswithampleincentiveto improvethe servicetheyofferto passengers.However,theirabilitytodo thislargelydependsuponpolicyreforms andavailabilityof technologies, which must address the economic factors drivingtheperceivedovercharging andgeneraldysfunction inthe auto-rickshawsector.These reformsneedtobebased onarecognitionofthedriver'sperspective,and thedaily reality of auto-rickshaw operation, which is one of long hours,highcostto revenueratios,andlimitedopportunities foreconomic mobility.Drivingan auto-rickshawprovidesa wagethat isunlikelytoplacethedriverandhisfamily above thepovertyline. Restrictive permitregimes, permit price inflation, and thenecessityof purchasing permits on the blackmarketmakeowning anauto-rickshaw impossibleor extremelydifficult,withmanypurchasesendingindefault andrepossession.

A platform that iseasyto understand for arickshaw driverwithalmostno formaleducationandconvenienttoa commuter whoisalways inahurrytogetarickshawis needed.Platformthatdoesnotburdenthe already limited economicmobilityofthedriver.Aplatformthatmakesuse

ofmoderndaytechnologiessuchasUPIandGPSforfinding customersaccordingtoconvenienceofthe driveraswellas finding rickshaws according to needs of the customer without charginghefty chargestoeither.

1. Inteli5processorandabove:Theprocessorneeded tobeusedshouldbeandIntel
2. I5processororhigherthanthatlevel.
3. 4GBRAM:TheRAMneededduringtheuseofthe projectshouldbegreaterthanorequal to4 GB.
4. 50 GB hard disk: The hard disk required for completeusageofprojectshouldbemorethan50 GB
5. AndroidMobileDevice:Anandroidmobilesystem with functioning GPSsystem and Playstore API level29orabove

B. Softwarerequirements

1. Android Studio 3.5.2: Android Studio is the integrated development environment for Google'sAndroidoperating systemdesigned specificallyforAndroiddevelopment.
2. Java virtual machine (JVM): JVM is a virtual machinethat enablesa computer torun Java programs aswellasprogramswritteninother languagesthatare compiledtoJavabytecode.
3. JRE: 1.8.0_202-release-1483-b03amd64
4. OpenJDK64-Bit ServerVMbyJetBrains
5. Windows Operating System: Windows operating systemistheOS platformrequiredforthe deployment oftheproject.
6. Google Firebase: Firebase is a Backend-as-a-Service —BaaS,app-developmentplatformon Google Cloud Platform. Primarily used for cloudcomputing,database management,user authenticationandhosting.
7. Google Realtime Database: Realtime Database is Firebase'soriginaldatabaseused forrealtime updatesonusers andrickshawdrivers

X. TECHNOLOGY

A. Java

Javais ageneral-purposeprogramming languagethat is class-based,object-oriented,anddesignedtohaveasfew implementationdependenciesaspossible. It isintended to letapplicationdeveloperswriteonce,runanywhere,meaning that compiledJavacodecanrunonallplatformsthatsupport Javawithouttheneedforrecompilation.Javaapplications aretypicallycompiledtobytecodethatcanrunonanyJavavirtual machine(JVM) regardless of the underlyingcomputer architecture. ThesyntaxofJavais similar toCandC++, but it has fewerlow-levelfacilities than either of them. As of 2019, Java was one of the mostpopular programming languages in useaccording toGitHub,particularly forclient-serverweb applications,withareported 9milliondevelopers.

JavawasoriginallydevelopedbyJamesGosling atSunMicrosystems(whichhassincebeenacquiredbyOracle)and released in 1995 as a core component of Sun Microsystems'Java platform. The original andreferenceimplementationJava compilers, virtual machines, andclasslibrarieswereoriginallyreleased bySun underproprietarylicenses.AsofMay 2007,incompliancewiththe specifications of theJava Community Process, Sun hadrelicensedmostofitsJavatechnologiesundertheGNU Pluginforapplets).

On,releasedonSeptember25,2018;Oraclereleasedfor thelegacyJAva8LTSthelastfreepublicupdateinJanuary 2019forcommercialuse,while itwillotherwisestillsupport Java 8withpublicupdatesforpersonaluseuptoatleast December2020.Oracle(andothers)highly recommend uninstallingolderversionsofJavabecauseofseriousrisks duetounresolvedsecurityissues.SinceJava9(and10and 12)isno longersupported,Oracleadvisesitusersto immediatelytransitiontoJava 11(Java13isalsoa non-LTS option).

General Public License. Meanwhile, others have developed alternative implementations of these Sun technologies, such as the GNU Compiler for Java (bytecode compiler), GNU
Class path (standard libraries), and Iced Tea-Web (browser)The latest versions are Java 13, released in September 2019,
and Java 11, a currently supported long-term support (LTS)

IX.REQUIREMENTSPECIFICATIONS

A. HardwareRequirements

XI. CONCLUSION

Thisundertakinghadafundamentalobjectivetobuildup a model capable ofhandling tremendous influx of client bookings inrealtimeaswellasstoringthe dataforfurther analysis. It wasconceivabletopresumethatthelevel of achievementofthe cinematographicincomeexpectationis veryunique dependentonthe typologyoftheneedyvariable utilizedintheinvestigation.The exactmodelshowedgreat measurableoutcomeswhenthe reliantvariablewastwofold and interim. Nonetheless, in what respects themulticlass expectation, theoutcomes wereexceptionallyalong way fromthe realworld,contrarilyaffectingthemodel.Wehave studied variousresearchpapersrelatedtorecognitionmodel.

Wehaveoptimizedthe Euclideandistancealgorithmas per our requirement to achieve better efficiency and accuracy.Inthis

report,anintelligentsystemwasdeveloped basedonmachinelearning andfeatureextractiontodeal with theproblem ofrickshawactivity.

Basedonwhatwehaveresearched,improvementscan be made for the existing systems to support real time bookingswithlessresponsetime.

Weaimedoncreatingadesperatelyneededintherealm ofrickshawsinMumbai,this modelcanbefurther extrapolatedtoothermetropolitancitieslikeDelhi and Bangaloreas rickshawsare deployedineverylargecityin India thereforeleavingascopeforlargeexpansion.Witha businessmodel tobolsterthisarrangement,anall-round product canbebuilt.

XII.ACKNOWLEDGMENT

Wesincerely thankourguideMrs.JesalkumariVaroliafor herguidanceandconstantsupportandalsoforthestickto ourbacks.Wealsothanktheprojectcoordinatorsfor arranging thenecessaryfacilitiestocarry outtheproject work.
We thanktheHOD,Dr. SheetalRathi,DeanAcademic,Dr.
R.R Sedamkar,thePrincipal,Dr.B. K. Mishraandthe collegemanagementfortheirsupport.

References

[1]SimonHarding,MadhavG. Badami,ConorC.O. Reynolds,Milind Kandlikar.*Auto-rickshawsinIndian cities:Publicperceptionsandoperational realities.*⟨http://dx.doi.org/10.1016/j.tranpol.2016.07.013⟩.

[2]EMBARQIndia,2011. SustainableUrbanTransport: TheRoleoftheAuto-rickshawSector.ResearchReport. WorldResourcesInstituteWashington,DC.

[3] Harding,S.,2011.A Few Questions about a Few Thousand Permits.Kafila. ⟨http:// kafila.org/2011/11/20/a- few-questions-about-a-few-thousand-new-auto-rick shaws- in-delhi-simon-harding/⟩.

[4] Harding,S.,2013.WhyAutorickshaws are Such a Pain for Everyone.August13th,. Yahoo!India. Available at:⟨https://in.news.yahoo.com/why-are-auto-rickshaws-such- a-pain-for-everyone–071219473.html⟩.

[5] Harding,S.,Hussain,A.,2010.On the Road to Nowhere? The System, Problems and Recommendations. Research

[7] Kleiner, Morris M., and Alan B. Krueger. 2013. “Analyzing the Extent and Influence of Occupational Licensing on the Labor Market.” Journal of Labor Economics 31 (2): S173–S202.
[8] Leah-Martin, Vincent. 2015. “When to Quit: Narrow Bracketing and Reference Dependence in Taxi Drivers.” Unpublished.

Report.AmanPanchayatReport,NewDelhi.
HindustanTimes,2012.Now,SMStoComplainaboutEr rant Auto-Rickshawdrivers.

[6]Hall,JonathanV., andAlanB.Krueger. 2015. “An AnalysisoftheLaborMarket forUber’sDriver-Partners in theUnitedStates.”PrincetonUniversity,Industrial RelationsSectionWorkingPaper587.

Design and DevelopmentAutomatic Gardening Systemusingweb API&Fuzzy logic

Mrs. Rashmi Thakur, Kunal Singh, Vignesh Singh, Pradyuman Sharma
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
rashmi.thakur@thakureducation.org, kunal4104@gmail.com, vigdrusingh@gmail.com, pradys.sharma10@gmail.com

Abstract—**theamountofwaterrequiredby aplant depends upon number of factors, which include plant'sspecies,soiltype,weatherand variousother microparameters.In general,bestwaytowatermost plantsisby applyingenoughto moistentheplant's entirerootsystem,and then lettingthesoildryout slightlybeforewateringagain.A soilmoisturesensor canbeusedtocheckthesoilnearplantroots andwater them accordingly.But,plantscanreceivewater naturally through rainandaswe know, toomuch water canbeasharmful astoolittle water tothe plants. Inthisprojectweareplanningtobuildan automatic wateringsystemwhichnot onlymeasure soilmoisturebutalso check theweatherdependingon whichitmakesdecisionto watertheplant.Tocheck theweatherwewillbeusingawebbased API which willcommunicatewithourAdrianomicrocontroller. If,in nearfuturethereischanceofrainfallthesystem willrefrainfromwateringthe plantsregardlessofsoil moisturestatus.Thisprojectwillhelp reducethe humaneffortrequiredandalso decreasewater consumption,whilemaintaining the healthof the plants.**

I. INTRODUCTION

Nowadaysthereis scarcityofwateralloverthe world, especiallyinIndiadue tolowgroundwatertableand irregularrainfallmostofthewatergetswastedin drainage.Thusthereisa needofmoreefficientand cheapalternativetotraditionalformofwateringthe plants.

Automatedplantwateringsystemestimateandmeasure theexistingplantandthensuppliesdesiredamountof waterneededbythatplant.Itis minimizingtheexcess wateruseaswellas keepingplantshealthy.Thissystem Savestimeandis accurate.Atplaceslikesocietiesand Buildinggardenswecan implementan automatic wateringsystemwhich is easyandaccurateforthe growthofplants.Thiswillminimizethehuman interventionandwillalso savetime.

To demonstratebuildingandenhancingofexisting automaticplantwateringsystems,weexploitavailable technologieswehavesuchas afreewebhostingand Freecloud todevelopsucha system.Webelievethatour Proposedsystemcanbeappliedtodifferenttypesof plants.

II. BACKGROUND

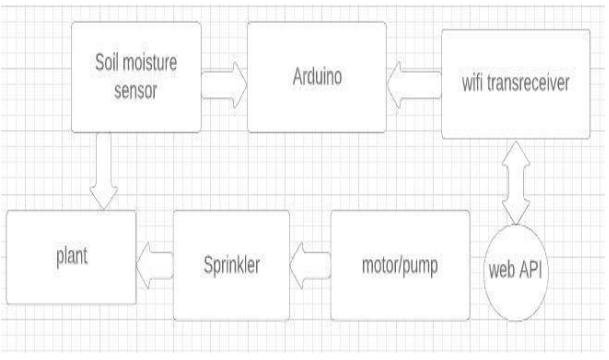
FuzzylogicNaturallanguage(likemostotheractivities inlifeandindeedtheuniverse)is noteasilytranslated Intotheabsolutetermsof0and1. (Whethereverythingis ultimately describableinbinarytermsis a philosophical Questionworthpursuing,but inpracticemuchdatawe mightwanttofeedacomputerisinsomestatein Betweenand so,frequently,aretheresultsofcomputing.) Itmayhelptoseefuzzylogicas thewayreasoning really worksandbinaryorBooleanlogicis simplya special caseofit.

InternetofThings(IoT)is a systemofinterrelated computingdevices,mechanicalanddigitalmachines, objects,animalsorpeoplethatareprovidedwithunique identifiers(UIDs)andtheabilitytotransferdataovera networkwithoutrequiringhuman-to-humanorhuman-to-computerinteraction

EmbeddedSystemiscombinationofcomputer hardware and softwarewhichis designedfora specific function.Industrialmachines,agriculturalandprocess industrydevices,automobiles,medicalequipment, cameras,householdappliances,airplanes,vending machinesandtoysaswellas mobiledevicesareall possiblelocationsforanembeddedsystem. Embedded systemtechnology reducesthecostand sizeofelectronic circuitforprojectdevelopments.

III. PROPOSEDIDEA

Figure1: Blockdiagram



Ourprojectis basedonthreefunctionalmodel.First functionis tomeasurethemoisturecontent insoil. Secondis tokeeptrackofweatherconditionsbytheuse of realtimedatafromopensourceWeatherAPI.Lastly analysesboththedata,comparetopre-setvaluesforthe givenplantanddecidewhethertostartthemotorfor watersupply.Thecompletesystemiscontrolledby Adrianoboardwhichis programmedusingAdrianoIDE software.

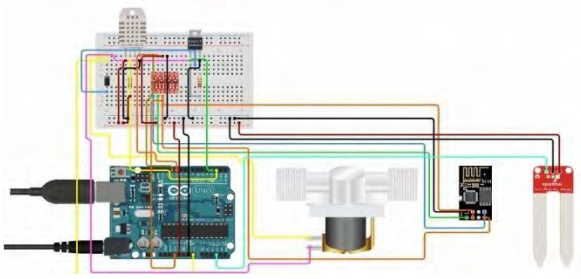


Figure2: PictorialRepresentationof System

A.Detectingsoilmoisturecontent:
Themeasurementofwatercontent inthe soilis achieved byusageof soilmoisturesensorwhichsendsan alertto themicrocontrollerwhen the soilmoisturedropsbelow Thresholdlevel.

B.GatheringWeatherdata
To tracktheweatherconditionsofnearfutureweuse openweathermapAPI whichis an opensourceweather APIthatprovidesdata inJSON formattothe microcontrollerineveryten minutesusingWi-FiTransceiver.Thisdata includesmeantemperature, probabilityof shower,airhumidity andwindspeed.

Automaticwateringandrecordingdata
Basedon theanalysisandcomparisonofcollecteddata withcertainpresetrequirementsandfuzzylogic Algorithmthe systemdeterminestheappropriate conditionstostartthemotorandwatertheplants

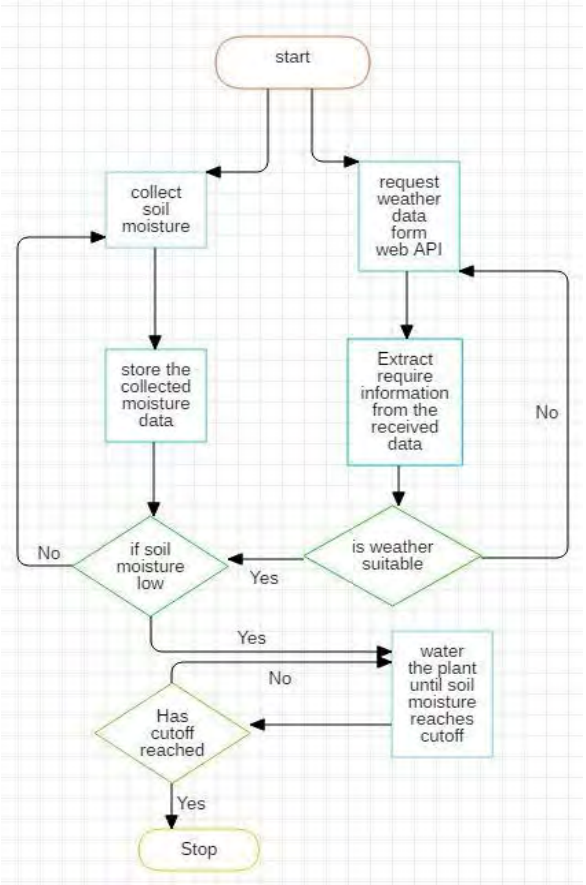
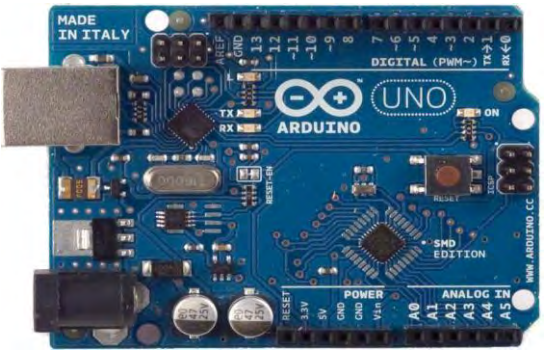


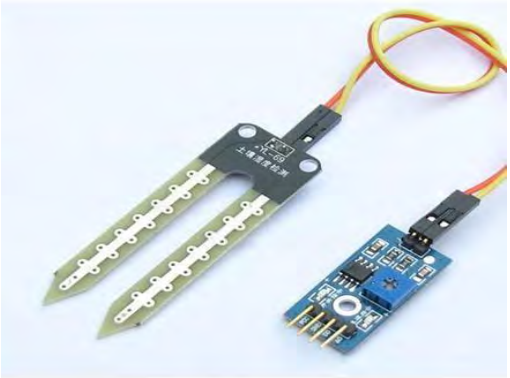
Figure3: FlowchartofAutomaticWateringSystem

IV. COMPONENTS USEDFORSYSTEM

1. Adriano Uno: This microprocessor is used to gathercollecteddatafromadjoining modulesandtakesdecisiontostartwatering.



2.Soil MoistureSensor:Soilmoisture sensorwillbe used to measuremoisturecontent inthesoil.



3. ESP8266-01 (Wi-Fi Transreciever):Tran'sreceiverwill be usedtocollectdatafromopenweathermap.org APItogetreal-timeweatherreports.

4. Relay: Themainoperationofthisdeviceis tomakeorbreakcontactwiththehelpofa signalwithoutanyhumaninvolvementinorder ToswitchitONorOFF.Wewillbeusingthisto Controlthemotortopumpwater.

5. Water pump:DCpoweredpumpsuse direct current frommotor, battery, or solar power to move fluid in a variety of ways.

6. SDcardmodule:Thismodulewillbe usedtocollectdata intheSDcardforfuture analysis.

V.RESULTANDANALYSIS

Duringtheprocessoftestingwehaveachievedpositive resultsfromthisproject.Itwasfound thatourplants wereabletogrowwithoutdehydrationwhilstusing limitedwater.Although,furtheroptimizationcan be Donetoourfuzzyruleset.ThedatastoredbySDmodule canbeusedtofurtheranalysesthepatternusingdata analysisalgorithms likelinearregression,K-mean clusteringetc.Wecan alsoaddaminicameramodule availablewithesp32Wi-Fimoduletostoregrowth processintheSDcardwhichcan beconvertedtoa time-lapsetofurtheryourunderstandingofthegardening.

V.CONCLUSION

Designanddevelopa smartsystemthatcansensethe contentofwater inthesoil.Andbyanalyzingtheresults irrigationbecomeseasy,accurateand practical.The Volumes, volumetricwatercontentsplayanimportant role inproducingtheoutput.

Theprimaryapplicationofthisprojectis to provideplantswithadequatewaterwithoutwastefuluse ofwaterandmanuallabor.Thisprojectis makingthe useoffuzzy inferencesystem todecidetheplant

Wateringprocess.Thisprojectcanfurtherbeextendedto largerscalecombiningwiththeprinciplesof rainwater harvesting;itcouldbeinstrumentalinareaswithsevere shortageofrainfall.

Theprinciplecan beextendedtocreatefully automatedgardensandfarmlands.Combinedwiththe principleofrainwaterharvesting,itcouldleadtohuge watersavingsifappliedintherightmanner.In agriculturallandswith severeshortageof rainfall,this modelcan be successfullyappliedtoachievegreatresults withmosttypesofsoil.

ThisSmartirrigationprovestobethesystem automatesforirrigationsystemand regulateswaterfor irrigationis donewithoutmanualUsingthissystem, solenoidvalvesandrelayboardcan becontrolled remotelywhichopenstheopportunitiestocontrolthe waterflowas wellastheelectricalflow.Irrigation systemis automatedwithdependson sensorReportthe pump is operatedbytheweatherconditionbysoil,rain andtemperatureconditionsthewaterpumpwillwork andbywirelesstrans receiverthedata is communicate andthe sensorreadingsareuploadedintocloudnetwork byWi-Fi___33technology

Implementing System to Detect and Isolate BlackHole Attackin WirelessSensor Network

Mr. Manish Rana, Abhay Singh, PrathamYadav
Department of Computer Engineering
Thakur college of Engineering and technology
Mumbai, India
Manish.rana@thakureducation.org, abhayyogendrasingh@gmail.com, ypratham4@gmail.com

Abstract—***AmobileAdhoc network(MANET)isa collectionofautonomousnodestoformanetworkwithoutuse of any supportinginfrastructure. The networktopology isunstructuredandnodes may enter or leave attheirwill.A node can communicate toothernodeswhichare within itstransmission range.Soanynodecanactasahost ortherouterinthenetwork, whichresultsinsecurityissuesinMANETs.Aprominent attack in MANETs isaBlackhole attack. In theblackhole attack, theattackernode broadcastsgoodpathstothenode falsely duringthe route-establishmentprocess.IftheBlackHoleNodeispresentin thenetwork,itwillreducethe network performancealongwiththe depletion ofthe energy inthe network.***

Thetechnique presentedis for detectionandisolation of blackholenodesfromthesensornetwork.Inthistechnique,the blackhole nodeisidentified bymonitoringthefakereplypackets that aretransmitted bythenodesanditwillberemovedfrom the network.Wevalidateour approach usingnetworksimulator

Keywords—
1.AODV:*Adhoc On-DemandDistance Vector*

2. MANET:*MobileAdHoc Network*

3.WSN:*Wireless Sensor Network*

4.NS2:*Network Simulator*

I. INTRODUCTION

MobileAd-hocNetwork(MANET)isacollection of autonomousnodestoformanetwork withoutuseofany supportinginfrastructure.Ifthetwocommunicatingnode sare withinthesamesensingrangetheycancommunicate toeach otherindividually, otherwisetheycancommunicatethrough multihopswherenodesactasintermediaterouters.Inthis typeofnetwork, nodescanenterinthenetwork orleavethe networkatanytimewithoutinformingto others.InMANETs, thereisnoguaranteethatapathfromsourcetodestination is free from malicious nodes. Due to lack of central coordination, thereareseveralattacksinMANETs.Active attackmayharmoralterthedatabeingtransmitted acrossthe network. IntheSybilattack,awell-knownattackisaBlack holeattack, whichiscreatedbyamaliciousnodebysendinga veryquickreplywithhighestdestinationsequencenumbe r andshortest path.Blackholenodecaneasilycorruptthe information. Toavoidsuchtypeofattacks,research community paysmuchattentiontowardsthesecurityof MANETs.
TheBlackholeattackisaddressedintheliterature either byconsideringthe energies of thenodes. The firstapproachdoesnotprovideaneffectivesolutiontothe BlackWholeattackas the maliciousnodescan beavailablein thenetworkwithdifferent energies. Inthesecondapproach, providingcertificationstothenodesthataremobileisdiffi cult.

II. MOTIVATION

AmobileAdhocnetwork (MANET) isacollection of autonomous nodestoformanetworkwithoutuseofany supporting infrastructure.Thenetworktopologyisunstructured andnodes mayenterorleaveattheirwill.Anode can communicate toothernodeswhicharewithinitstransmission range .So any node can act as a host or the router in the network,whichresultsinsecurityissuesinMANETs. A prominentattackinMANETsisa Blackholeattack.

ObjectivesofProject

Withtheadvanceintechnologyandmethodology moreand morenetworksarebeingdevelopedduetowhichthesecuri tyof thesenetworks hasbecomeamajorconcernandmanyexperts andthinkersinthefieldoftechnology havecomeupwithnew techniquesin order todefendtheraisingsecurityissues.

Scopeof theproject

Asthetechnology evolvesdaybydayvarious breachesinthe existingtechnology cometolightandhencevariousattacks whichshouldbeprevented, thisprojectparticularly focuseson identification andprevention ofBlackHoleAttack.InBlack holeattack,anadversaryannounces avalidshortestpathtothe destinationbytransmittinganomalous routinginformation.Out of

repliesfromdifferentnodesduringroutediscovery, thesourcenodeconsiderspathfromthemaliciousnodeco nsidering itasagenuine node having fresherpathtothedestination.Asa result,a bogusroutewillbecreatedthroughthatnode.

ProjectsImpactAnalysis

MANETsaremorevulnerable toattacksbecausetheyhave somespecificcharacteristics ascomplexityofwireless communication andlackofinfrastructure. Hencesecurityisan important requirement inmobileadhocnetworks.Oneofthe attacksagainst networkintegrityinMANETsistheBlackHole Attack.Hencesecurityisan importantrequirementin mobilead hocnetworks. Oneoftheattacksagainstnetworkintegrityin MANETsisthe BlackHoleAttack.

ExpectedOutcome

Nodes withoutintroductionofanynewcontrolpacket; defaultroutingpackets propagateinformation of adversariestoothernodesinthenetwork.

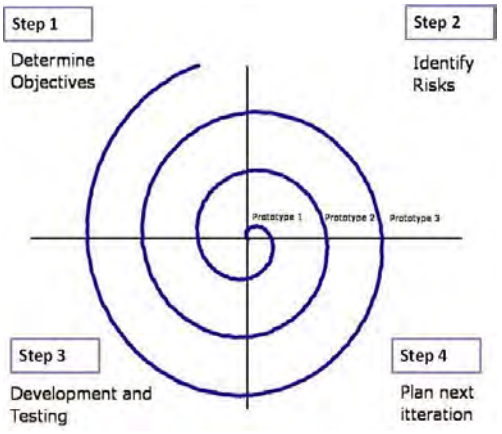
Themechanismprovideshighpacketdeliveryrate withnoticeable normalized routingoverheadand acceptable average ToovercomeBlackHoleAttackinWirelessSensorNetwo rk sothatwecanprovideasafenetworktocommunicate with otherswithoutthefearofspoofing oranychancesof denialof serviceattack.

Applicationofproject

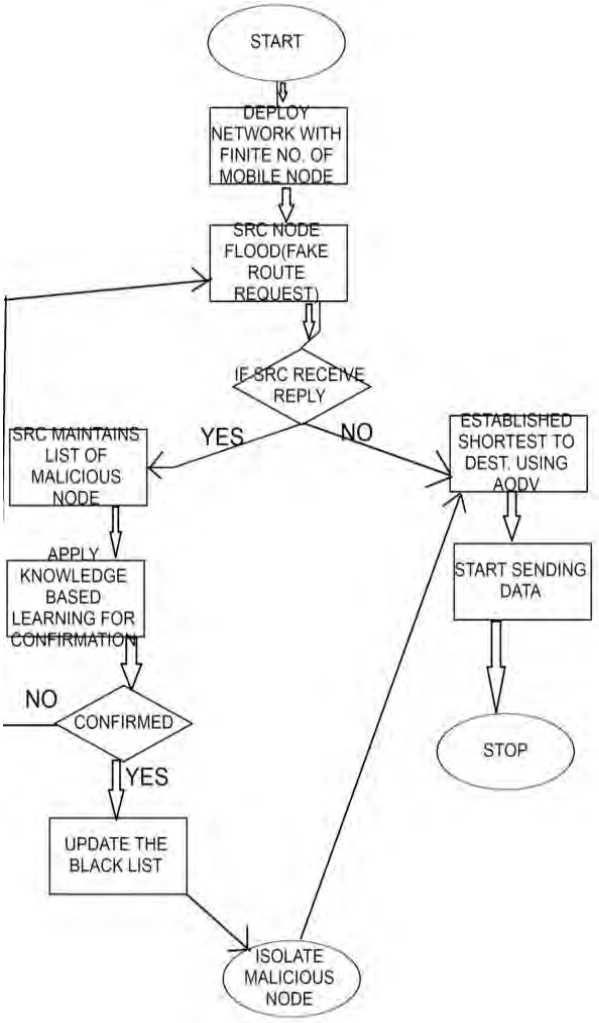
I. TodetectandIsolateblackholeattack Routingprotocol.
II. TheimprovementofroutingprotocolsinWSNin termsof security.
III. Analysisof RREQfloodingattackinMANET.

III. METHODOLOGY

TheProjectfollowsthesoftwaredevelopment lifecycleand usethespiralmodel.Thespiralmodelconsists ofthephases showninthefigure.



IV. FLOWCHART



V. CONCLUSIONS

• Cooperativetrustedenvironmentamongmobilenodes inMANET isabsolutelyvital.Inthisprojectwe provideimprovementin route discoveryprocessof AODV protocoltoisolatemultipleBlack hole and Gray-holenodes.

• AODVfailstoremovemaliciousnodesduringroute discovery processandthereforedoesto succeedto transferalldatapacketstothedestinationunderattack.

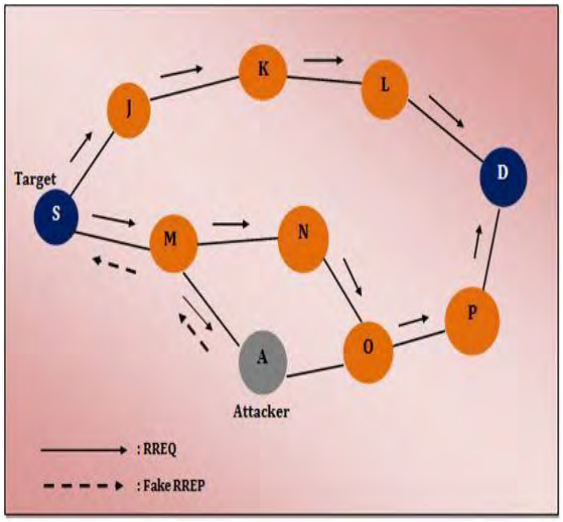
• Ontheotherhand,R-AODVprovidesasimpleand efficientwaytodetectandisolatemultiplemalicious

VI. ACKNOWLEDGEMENT

WewishtoexpresssincerethankstoourComputer departmentforarrangingsucha niceevent.Sothatwecan researchallthisaboutNetworking.Wealsoextendour Heartfeltthankstoourcolleagues,familymembersand well-wishers.

SystemAnalysis(functionalmodel,structural modeland behavioralmodel)

Itisaprocess ofcollectingandinterpretingfacts,identifying theproblems,anddecomposition ofasystemintoits components. System analysisisconductedforthepurposeof studying asystem oritspartsinordertoidentifyits objectives.Itisaproblemsolvingtechniquethatimproves



thesystemandensuresthatallthecomponentsofthesyste m workefficientlytoaccomplishtheirpurpose.

References

1.Baviskar B.R, Patil V.N," Black hole attacks mitigationand preventioninwirelesssensor network", International JournalofInnovativeResearchin AdvancedEngineering (IJIRAE),Volume1,Issue4, pp.167-169,May2014

2. MaanyounisAbdullah,GuiWeiHua,NaifAlsharabi, "Wireless SensorNetworks Misdirection Attacker ChallengesandSolutions",2008IEEE978-1-4244-2184-8/08/

3. Roshan Singh Sachan, Mohammad Wazid, D.P.Singh, Avita Kata and R.H. Goudar,"Misdirection AttackinWSN:Topological Analysisandan Algorithm for Delay and Throughput Prediction",2012IEEE978-1-4673-4603-0/12/

4. S.Anbuchelian,Selvamani.K,Chandarasekar.A"An EnergyEfficientMultipath RoutingSchemeby Preventing ThreatsinWireless SensorNetworks", ElectricalandComputerEngineering (CCECE),IEEE27thCanadianConference,2014

5. Kalpana Sharma and M K Ghose"WirelessSensor Networks:AnOverviewonits SecurityThreats"IJCA Special Issue on "Mobile Ad-hoc Networks" MANETs,

Askme: A Conversing Agent Of Indian Railway

Rohit Singh,Kishan Singh,Aishwarya Rane,Vidyadhari Singh
Computer Engineering Dept.
Thakur Collegeof Engineering and Technology
Mumbai, India
rssinghrohit1402@gmail.com,kishansingh180.ks@gmail.com,aishwaryarane5@gmail.com,vidya2686@gmail.com

Abstract— **Public Intercom is an integral part of the railway system. At many public places like railway stations, when people walk-in they look for the basic information and this task is still being carried out by humans and is person-specific. The inquiry window becomes a bottleneck and leads to bad user experience. We have proposed a "Digital Public Intercom" system that can be used to manage public profiles in any physical space using Wi-Fi or data network connection. It consists of a chatbot that can answer frequently asked questions for location and also various questions related to the trains. The proposed system will surely help to improve the user experience. An individual would be able to schedule their train trips with the help of this system.**

Keywords— Chatbot, NLP, Train Suggestions, PNR

I. INTRODUCTION

Indian railways are backbone of transportation system. Around 30 million people travel daily in India. Public Intercom is an integral part of Railway System. It comes into the picture when user queries are infinite and enquiry window becomes a bottleneck due to chaos. Indian trains are always late which wastes passengers time. Almost all of the people who travel by train prefer to have their trip properly scheduled to reach the destination on time. Many people check the IRCTC website to check their (Passenger Name Record) PNR status and if it is confirmed, they look for basic information like live status of the train, "On which platform the train will arrive?" etc. Even to obtain the above-mentioned information from a mobile phone application, the user has to download an application which consumes internet bandwidth and wastes time as well. Thus, the passengers can save their time and internet data by using the Chatbot by just typing questions and getting their queries solved at faster rate.

II. LITERATURE SURVEY

A. Study on the Public Information Service Systems of Railway Industry

Railway transport has a special important status in the development of economy society. Transportation costs have a significant influence on people's choice of travel. The above paper focuses on railway information service for public and the main factor is how to achieve effectively sharing of information resources, how easily and efficiently it can be transmitted. An aim is to provide satisfactory customer Information service considering all the conditions of existing technologies.

B. Android Based Educational Chatbot for Visually Impaired People

The above android application provides an educational based chatbot for physically impaired people. The visually impaired people ask queries through chatbot which gives response to all the education related questions. The application can be easily launched using google voice search. The output will be received in terms of voice using text to speech method. An application gives voice instructions for its usage, once opened. Responses from the chatbot will be voice based as well as text based. So that everyone can use the application.

C. Home Automation using IoT and a Chatbot using Natural Language Processing

The purpose of the mentioned web application is to control the electrical home appliances over the internet. It allows user to text information through chatbot in order to control the appliances. The sent information will be processed using Natural Language Processing and any device connected to the local area network of the house controls the other devices. The NLP techniques like tokenization, removing stop words are performed. This system uses a Wi-Fi based Local Area Network Protocol. An interference problem can be solved by giving a specific Local Area Network to each application. Messages for controlling devices can be sent by a chatbot.

D. A Platform for Human-Chatbot Interaction Using Python

This paper describes a chatbot agent which is capable of communicating using natural languages. Also, it explains how human-chatbot interactions are analysed by providing web connectivity to evaluate chatbot on web-based platform. As users are having confidential problems in sharing their own data online, a web-based platform is provided to users in order to communicate with the world by their chatbot easily using python.

III. REQUIREMENT

The requirement for this project includes some software specifications which are: -

1. Heroku server- This is an open server which can be used to create the link between the chatbot and railway API.

2. Heroku CLI & Git – This is a command line interface for working and deploying the local repository to the Heroku server which take the help of Git or Github desktop.

3. Django- It is an open source high level python web application framework.

4. MYSQL- an oracle backed open source Relational Database System (RDBMS).

5. Railway API- Responsible for providing the information as JSON object like PNR status, Train between the stations, running status, etc. The module uses a unique API key provided by API providers such as erail.in and railwayapi.com.

6. NLP- Natural Language Processing is concerned with how technology can meaningfully interpret and act on human language inputs.

IV. SYSTEM ARCHITECTURE

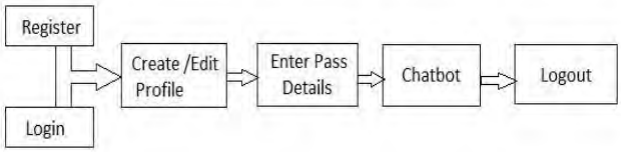


Figure.1. System Architecture

1. Input- To get PNR details user can give input as: "PNR 85451234".
2. Output- The output of above query will be "TRAIN NAME-Siddheshwar Express DOJ-1/2/2020, TRAIN NO. -13552, CUR STAT-RLWL/021‖".
3. Backend- a backend that is used to connect web application and railway API is MYSQL.

3.1 Heroku server- This is an open server which is used to create an application to store the backend database and creating the link between the chatbot and railway API.

3.2 Python- The backend is programmed using Python version 3.7 and pip. This is used to fetch the JSON object from the Railway API.

V. PROPOSED SYSTEM

Profile management helps user to store their data .User can enter source and destination to fetch list of trains to reach particular station by time. Alert notification regarding expiry of pass and PNR fetching facility are parts of public intercom platform. It also includes a chatbot that will answer any frequently asked queries by user. The queries could be of any type for example a query regarding particular location or a query like" On which platform a particular train will arrive?" and so on.

VI. IMPLEMENTATION

1. Profile Creation - Provide login panel for users to store their data.
2. Integration - Integrate Railway API to web application.
3. Storage - Create back end to store user profiles and pass data.
4. Alert Notification – Create a process to send push notification 2-3 days before the specified date.(pass expiry date)
5. Query Evaluation - Specify the intent name and describe the expressions that a user can ask. Specify the parameter and create a logic adapter.

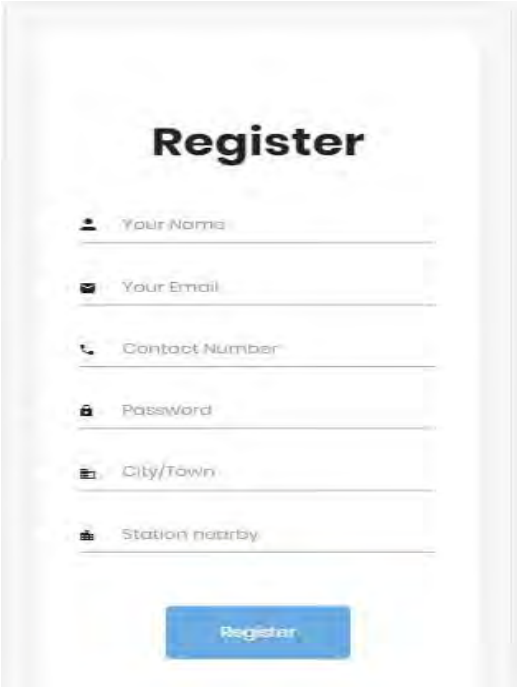


Figure.2. Registration Portal

Figure.3. Pass Input

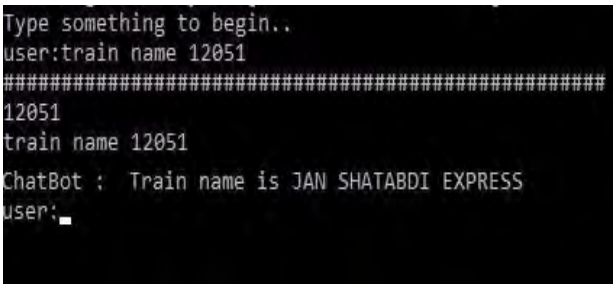


Figure.4. fetching train name from train number

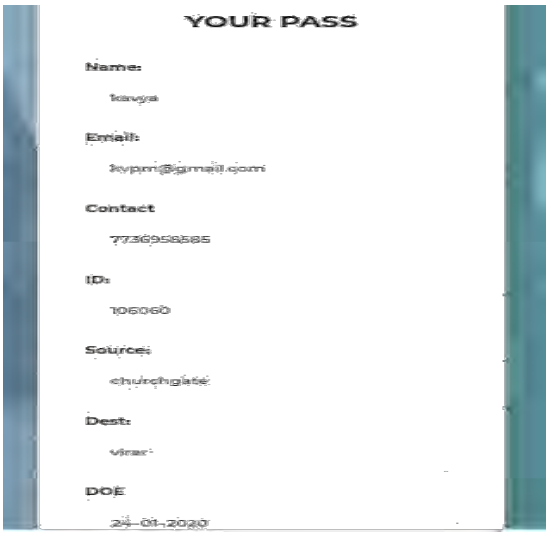


Figure.5. fetching train live status

The system will send notifications 2-3 days before user's pass expire.

VII. CONCLUSION

This paper successfully explains and implemented the public profile management and chatbot which can be used to get the required information from Indian Railway. In the future, we will include the voice automated chatbot for handling queries. The key feature of an effective chatbot is to make sure it enhances your customer's experience. This chatbot will benefit the passengers by saving time, memory space.

VIII. ACKNOWLEDGMENT

We want to thank our guide Mrs. Vidyadhari Singh., Assistant Professor, Thakur College of Engineering and Technology, Mumbai, India for his continuous support and his encouraging and fruitful advises in accomplishing this task.

References

[1] Naveen Kumar M, Venkatesh Prasad A, Linga Chandar P C, Sumangali K- ―Android Based Educational Chatbot for Visually Impaired People‖, Issue, 2016.

[2] Guo Chunfang , Wu Hao, Sun Zhongyue Guan Zhongliang-―Study on the Public Information Service Systems of Railway Industry‖, Issue, 2010.

[3] Anirudh Khanna, Bishwajeet Pandey, Kushagra Vashishta, Kartik Kalia , Bhale Pradeepkumar and Teerath Das-―A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence, Issue ,2015.

[4] Cyril Joe Baby, Faizan Ayyub Khan, Swathi J. N.- ― Home Automation using IoT and a Chatbot using Natural Language Processing‖, Issue , 2017.

[5] Bhaumik Kohli, Tanupriya Choudhury, Shilpi Sharma, Praveen Kumar-―A Platform for Human-Chatbot Interaction, Issue, 2018

BlockChainBasedResumeVerification System

Dr. Harshali Patil, Rushikesh Mokashi, Suchit Gupta, Srihari Nair
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
harshali.patil@thakureducation.org, rushimok73@gmail.com, suchitgoyal000@gmail.com, sriharinair1998@gmail.com

***Abstract* - BlockchainBasedResume Verification Systemisa systemwhich willhavethe primary goalof verifyingandconfirming the educationalandprofessional qualificationsof everyindividual. This systemwillbeimplementedonEthereum asourblockchain,Word PressforourCMS, mongoDBasour database andNodeJS, phpfor serverscripting.Also,a scoring system will be added onour website, based on asystem ofreps and nags discussed later.[2]Previousapplications which have implementedsomething similarto this software isLinkedIn.Butanareawhichit lackedinwasthe verificationof the qualifications.[1]We believe thatwithour systemwe couldovercomethishurdleand fulfilour main objective.[1]Our mainobjective being-tocombat people lyingontheirresumeabouttheir qualifications. [1] Weaim to createa trustworthysystemandthuseffectivelyend this problem.**

Keywords: Block chain,Ethereum, Resume, Verification, WordPress, NodeJS, Reps, Nags, Score, mongo DB

I. INTRODUCTION

Fora longtime,people have beenlyingon theirresumestogetbetteropportunitiesand theyhavelargelybeenabletogetawaywith itbecausecurrently,there isnowayto verifytheclaimsmade ontheresume.This isabigproblemfor corporationsand educationalinstitutesfor hiring'sand admissions,andwe aimtosolvethis problem.Also,there presentlyexistsno resumescoringsystem,asitishardtoputa valueona largelysubjective approachto analyzingresumes.The systemof repsand nags[2] weaimtoimplementwillsolve thisproblem,andalso make thetaskof Hiring departmentsinmany organizations much easier.

A.Block chain

Blockchainisa databasethatisdistributed andmaintainsrecordsof acontinuously-growinglistsecured fromtampering and revision. Each blockisconnected tothe previousblock viaa linkwhichhasa timestampattachedonit.[5] The main technical innovation ofbitcoin-which was conceivedin2008 andfirstimplementedin 2009-Servesasthepublicledgerforbitcoin transactions.New blockscanbecreatedby every user, theyare alsoallowedtosend newtransactionsandattempttoverifythose transactions,andcreatenewblocks.TheInspirationfor other applicationshasbeen the bitcoinblockchain.

B.Ethereum

Ethereumisusedfor decentralized applicationsor dAppsandisbasically blockchain-based platform. An Ethereum Virtual Machine orEVM isa decentralized virtualmachineprovidedby Ethereum,which can execute scriptsusing aninternationalnetworkof public nodes. [6] In contrast to the Bitcoin Script, the virtualmachine's instructionset,isthought to beTuring-complete. An internal transaction pricingmechanism, known as "Gas", is usedto mitigatespam and allocate resources on thenetwork.

C.WordPress

WordPressbasedonPHPandMySQLis a contentmanagementsystemthatismostly usedwiththe SQLite databaseenginebut can also be used by MySQL orMariaDB database servers. It has a template systemanda pluginarchitecture, known asthemesinsideWordPress. Bloggingissomething whichismost associatedwithWordPressbutisnowoften usedtosupportother typesof webcontent aswellincludingmoretraditional mailing listsand forums&onlinestores.

D. MongoDB

MongoDBisadatabase programwhich provides cross-platformdocument-oriented database,itusesJSON-likedocumentswith schemaclassifiedasaNoSQL database program.

E.SmartContracts

A smartcontractissuchthatitisusedto digitallyverifyandprotectandenforce the termsunder which the contracthad been initiallywritten.These allow the performance ofcredibletransactions withoutthird partiesor anyexternal participation. [6]Thesetransactions are trackableandirreversible. Proponentsof smartcontractsmaybemade partiallyor fullyself-executing,self-enforcing,orboth,a claimthatmany kindsof contractual clausesmake.The mainpurpose of a smart contractsistoprovidea guaranteeand protection,thatissuperior totraditional contractlawandtoreduceother transaction costs associatedwithcontracting.Mostof these smartcontractsareusedin implementationof varioustypes of smart contracts.

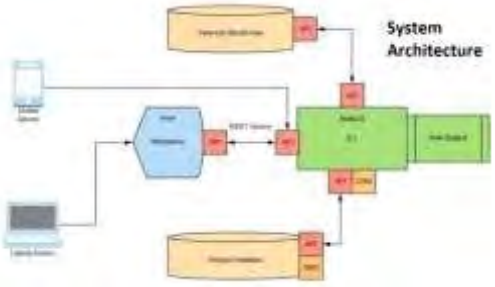
II. LITERATURE SURVEY

We have surveyedseveralresourcesand papers.Fromthese paperswe understood several principlesanddida gapanalysis. WefirstlookedatMethodsandsystemsfor employment and education verification usingblockchainweunderstoodhowtouse blockchaintodeployverification system butwe feltthatit hadalackofexplanation abouthow technologystackmay affectthe effectiveness.Further,we tookalookat multiple papers regardingComputerized methodandsystemforcreatinga digital resume,[4]A systemandmethodfor creatinga databaseof independently validated educationaland work experience profiles(curricula vitae)usingblockchain smart contracts,feedback based reputation ontopof thebitcoinblockchain,A Study on PrivilegeElevation Attack Management for SmartTransaction Securityon BlockchainEthereumBasedSystem.Agap analysiswasconductedonallof theseand taken into consideration for our paper. [5]

III. METHODOLOGY

A.ResumeCreation

Tocreatearesumeon ourwebsite,auserhas tofirstcreateanaccount on our websiteand select h i s a c c o u n t t y p e w h i c h w i l l b e a n ‘Employee’oran‘employer’.Thesewillbethe tworolesonourwebsite. Anyclaimsmadeby theemployee willbe verifiedby the organization orhispeers/supervisor. Upon loggingin,theuserhasto completearesume form,thedatafromwhich willbestoredinto themongoDBdatabase.



B. Database Storage

Themongo DBdatabasewillstorelogin credentials, resumedetails,userdetails,score and multipleotherdatafields. Thedatafrom themongoDBserverwill beinsertedtoand extracted fromusingREST APIservice.Weare intended on using RESTAPIaswearehopeful of further developingamobileapplicationand expandforward fromourwebsite.[3]

C.ServerScripting

WeareusingNodeJSand PHPforourserver-sidescripting.Themajor useofNodeJSisto fetchandpost to ourAPI’sfrommongo DB. NodeJSwillbe mainlyusedfordata transmission.[3] Also,PHPisbeingusedasit worksverywellwithWordPress,asWordPress itselfiscodedin PHP.Thus, PHPallowsusto implement avariety ofcustomfeatures on WordPress, as WordPress doesn’t match all ourneeds. WordPresshas avariety ofplugins wecanuse,buthoweveralot ofadditional programmingisneededthroughPHPto make them worktoour needs.

D.APICalls:

Everyinteraction withthedatabasewill happenthroughaREST APIcall.Thisis implemented keeping in mindfuture expansion ofthewebsitetootherplatforms andcreatea modularapproach.Currentlya fetchand astoragecallwillbemadetothe database.This mightincreaselateras development continues.

E.ScoreSystem

Everyprofilewill havea correspondingscore associatedwithit.[1] Thisscorecanbe changed via repsand negs. A rep isapositive increaseintheusersscore,whileanegisaDecreaseinthesc ore.Wehavereps andnegs [2]tied toeventsinourwebsitethatshall trigger them. These havebeendiscussed indepth furtherlater.

F. Verification using Block chain Every verifiedclaimon thewebsiteand the particularchangeinscore tiedtoitshallbe updatedontheblock chain.[7]Wewillbeusing Ethereumas ourblockchainandwillbeusing Ethereum smart contractsto reflect the changesmadetotherepscoreandverification statusof aclaimontothe Blockchain.[1]

IV. CONCLUSION

Thus,this projectwill beimplementedusing Block chaintechnologyand theusage of smart contracts.Asystemofrepsandnegs[2]willbe implementedasasystemofpointsin the website toincentivize website activityand enableaprofilescoring system. Every transactionwillcausetheexchangeofrepsand negs [2]tobeupdated inthepersonsresume. A verifiedresumewillbecreated and reputationwillbeused togradeeach resume.

References

[1]. Jaspreet Randhawa, “*Methods and systemsfor employmentandeducation verification using blockchain*”,lens.org, Oct2018

[2]. Davide Carbon, “*Feedback Based Reputationon TopoftheBitcoin Blockchain*”, arxiv.org, Feb2015

[3]. Guy Zyskind,Oz Nathan,Alex 'Sandy' Pentland “*Decentralizing Privacy:Using BlockchaintoProtectPersonal Data*”, IEEE,May2015

[4]. Daniel Jacobus deWaal, “*Computerized method andsystemforcreating adigital resume*”,patents.google.com, Nov 2012

[5]. GaryMcKayAdiBEN-ARIBrianMcnulty, “*Asystemand method forcreating a databaseof independently validated educational andworkexperience profiles (curriculavitae)using blockchainsmart contract*”, patents.google.com, March2016

[6]. Mayukh Mukhopadhyay, “*EthereumSmartContractDevelopment*”, Feb2018

[7]. Chris Dannen, "Introducing Ethereum and Solidity", March 2017
[8]. Min, Youn ,"A Study on Privilege Elevation Attack Management for Smart Transaction Security on Blockchain Etherium Based System", koreascience.kr, Feb 2020

Online Web-Based Assessment Platform

Vikas Sing, Mithil Bhuta, Zaid Khan, Yash Mehta
Department of Computer Engineering
Thakur College of Engineering & Technology
Mumbai, India
mithilbhuta@gmail.com, zaidkhan291@gmail.com,yashmehtayj@gmail.com, vikas.singh@thakureducation.org

Abstract— **Our OnlineWeb–Assessmentstageisan on-line appraisalstageto leadevaluationseffectivelyoverthenetand beatingthecons identified withthetypical manualpen-paper based tests.Weexpecttoexecuteawebelectronicappraisalstage forteachers.Theseendeavorswillendin constantimprovement andincrementtheusageofthePCsinsidethestage.Thisweb baseonlineappraisalstagewillsupplanttheoldpen andpaper group bysigning on.It'll bemore secureandlessdefenseless against assessment mistakes and incomplete assessment as depletedamanualway.The inspector'sactivitygets simplerby gettingtheassessmentprocedureandin thismannertheoutcome ageprocessrobotized.**

Keywords-online examination; algorithm; security; teacher;Evaluation;results.

I. INTRODUCTION

OurOnlineWeb –Assessmentstage isanonlineappraisal stageto directevaluationsproficientlyovertheweb,defeating thecons relatedwiththecustomarymanualpen-paperbased tests.We expect toexecuteanonlineelectronic appraisalstage forinstructors.Theseendeavorswillbring aboutceaseless improvementand incrementtheutilizationofthePCsinthe stage.Thiswebbaseonlineappraisalstagewillsupplantthe old penandpaper designbygoingon theweb. Itwillbe increasinglysecureandlessinclinedtoassessmentblunders and incompleteassessmentasdoneinamanualmanner.The inspector'sactivitygetssimplerbygettingthe assessment procedureandtheoutcome age processrobotized.It'llfinishthe need of assessing the researcher effectively through a completely computerizedwebappraisalstage thatsparestime, givesquickandexactoutcomes,andso onit'llbesimplerto make newtests,giveresultsandmuchfurtherimprovethenet evaluation stage bytaking fittinginputsbythetest takers. Online assessmentframeworkhelps understudies tosupplya quick anddirectgratitude toshowupforthetest. It additionallygivestheoutcomesfollowingtheassessmentwit h 100%exactnessandsecurity.Understudy canentertoperform testjustwiththeirsubstantialusername andsecretphrase.This assessment contains different decisionbasedinquiriesandpropernumberofalternatives.

A. Objectives

Therearenumerousgoalsoftheventurewhichfacilitates theentireprocedurefortheanalystaswellasforthe testtaker. Itdiminishesthe remainingtaskathandofthewebevaluation analystbytakingtheentireprocedureontheweb.

- Integratevarioustestsandstayaway fromthedangers relatedwithmanual conductionandchecking.

- LocationIndependent:Studentsand analystsneednot travellongseparationsonlyforthe testandrathersavemoney onscheduleandcostsbydirectingtheentireprocedurewith out breakingasweatof sittingathomeoran appraisalplaceinside nearnesserenely.

- It willgivea substantially moresolidandquicker outcome arrangementfortheunderstudies.

- Provision for test and arrangement keys for evaluations towardthefinishoftheappraisals.

- Features whichthemanualpenpaperbased routeneglectstogivelikeexactplannedtests.

- Itstaysawayfromthe dangerofmanualblunders like fractionalassessmentorremovalofarrangementsheets.

- Randomizationoftheinquiriesforeveryunderstudy whichisn'tregularlygiveninpenpaperbasedappraisalstake n.
- Loginhighlightto guaranteethesecurityoftheentire procedurewhere the understudiesandanalystshavevarious degreesofconfirmationauthorityandaccessibility.[1]

B. Project Scope

Withafewonlineappraisalsandassessmentsbeingtaken normallybyagreatmanyunderstudies,thiswebevaluation stagehasanexceptionallywideextension.Itparticularlyhas a moreextensivedegreeandapproachcontrastedwiththe manualpen–paperway.This canbeutilizednot exclusively toevaluateunderstudiesbyorganizationsyetinadditioninth e businessworldtosurveyworkers.
- It additionallyspares theclientfrommakinga trip longseparationstogiveatest.
- Thesearestageswhichlettheunderstudiestakethe testfromthesolacesoftheir home.
- willempowerthesynchronousconductionoftests allthroughthenation,andsoforth.
- Itwillgiveacompletestage totheunderstudiesto assessthemselvesbyproducingresults andreports.

• Thereisnorequirementfortheinspectortobe availableduringtheconductionofthetest.
• Provisionofbettersecurityandstraightforwardness whencontrastedwiththedifferentdangersassociatedwiththe conventionalmanualcheckingtechnique.
• Designed toencourageadministratorclientsand understudyclients.

C. ProjectsImpact Analysis

It diminishes the outstanding task at hand of theweb evaluationanalystby takingtheentireprocedureontheweb. Theframeworkactualizedcouldfillinasa heartytestingstage forcontenderfortarget questions. (variousdecisionquestions) Through thisproposed frameworktoo, understudiescanbe reviewedallthe moredecently andimpartiallygiventhatour methodology executesacatchphrasematchoftheunderstudies accommodationagainstthecheckingguidesetby the inspector.[2] Consumption on paper is diminished subsequently sparingexpensetotheorganizationorthe foundationandfurthermoreensuring theenvironment.Time sparedfromutilizingthe frameworkcanbeutilizedfor progressivelyprofitableexercises.

II. SYSTEMDESIGN

A. SystemArchitechture

Theexecutivemoduleand understudymodulesincorporate theirpieceofcapacities totheassessment.Themanager includesenlisteddata, alterorerasesit asrequired.The inquirynumbers areconsequentlyproduced.The totalsubject astuteconsequencesofunderstudiescanbeseenbythetestov erseerwheneverafterthefinishingthetest.[3]The understudyjustshouldlogintogotothetestandinthewake offinishing andpresentingthetesttheoutcomeisinasplit secondcreatedthe assessmentframework.Theelement"Understudy"cangive testaftertheindividualinquestionaccessestheframework. Thesubstance"Director" cantransferinquiriestoberepliedbyunderstudyintotheasse ssmentdatabaseutilizingquestiongroup,set thetestguidelinesanddesigntheright choicesorsetofalternativesforthe testquestions

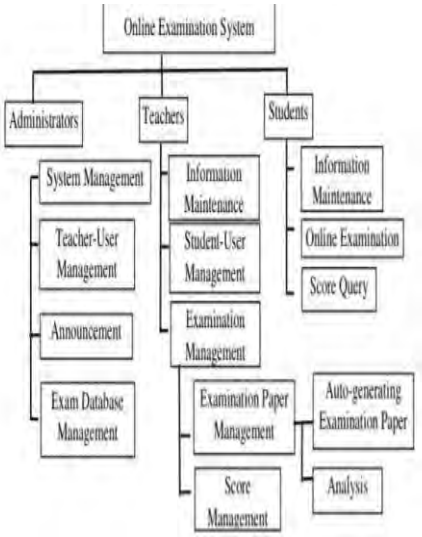


Figure1:SystemArchitecture

B. DataFlowDiagram

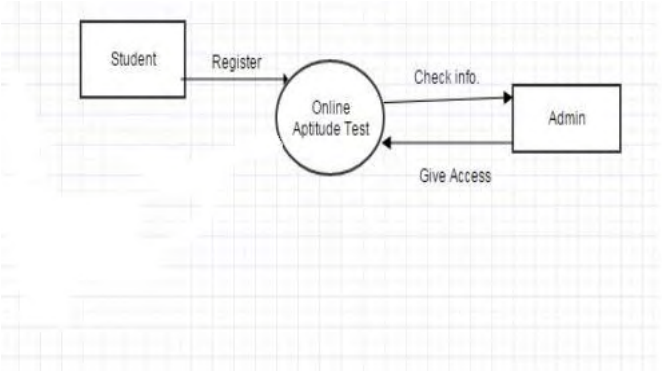


Figure2:DataFlowDiagram

Aninformationstreamgraph(DFD)utilizesfew crudeimages tospeaktotheusefulnessperformedbythetaskandthe streaminformationamongthevariouselements ofthe undertaking.Theinformationstreamchart delineatedinfigure Bunderneathshows therelationship amongtheelements in

.

C. StructuralModel

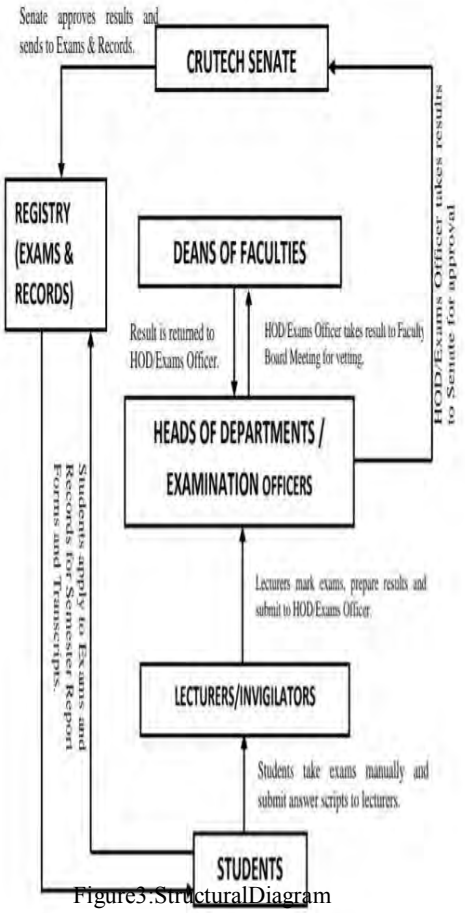


Figure3:StructuralDiagram

Structure outlineisareasonable displayingdevice used toreportthevariousstructuresthatmakeup aframework,for example, a database or an application. It shows the progressionorstructureofthevarioussegmentsormodules of theframeworkandshowshowtheyinterfaceandconnectwith oneanother.Itisadeviceusedtomanageengineers to guaranteethatallpiecesof theframeworkfunctionasplanned comparabletothevariousparts.

Structureoutlineis an improvementdeviceutilized in displayingthevariouspiecesofaframework,fromthe review onhow the individualpartsconnecttomaketheentire,to demonstratingthesubtletiesofthe littlestpartsthemselves,for Example, thevariousarticlesandclassesbeingutilized in programmingtheframework.

Astructuregraphimagineshowaframeworkfunctionsfrom theunderlyingcontribution,topreparingand,atlonglast,to the idealyield. It isparticularlyvaluableindecidingthe Entiretyoftheinterfacesrequiredbetweenthevariousparts andassistsengineerswithconcurringonhoweachpartought tobeassociateddependentonthemodelsbeingappearedon thestructuregraph.

D. UMLDiagram

Use cases give an elevated level perspective on the framework.UseCasesdemonstratingis acompellingmethods forspeakingwithclientsanddifferentpartners aboutthe framework and what isplanned to do. An utilization case portraysgroupingsofactivitiesaframeworkplaysoutthat yieldadetectableaftereffectof significantworthtoaspecific on-screen character.[4] Itbolstersnecessities building exercises and theprerequisite procedure that catch what a frameworkshould do,essentially,frameworksutilitarian necessities.

This is anutilizationcasegraphforonline assessment framework.Theutilization casegraphshowstwoon-screen characters-analyst andunderstudy, andhowtheyconnect withtheassessmentframeworktoaccomplishwhattheyneed, whichincorporatesplanningquestionback,getting ready assessment,takingtestandcheckingontestresult.[5]There are incorporated use cases that model a consideration of conductofferedbyanotherutilizationcase,andbroadeneduse casesthat modelthe excellentconductthroughoutbase use cases.

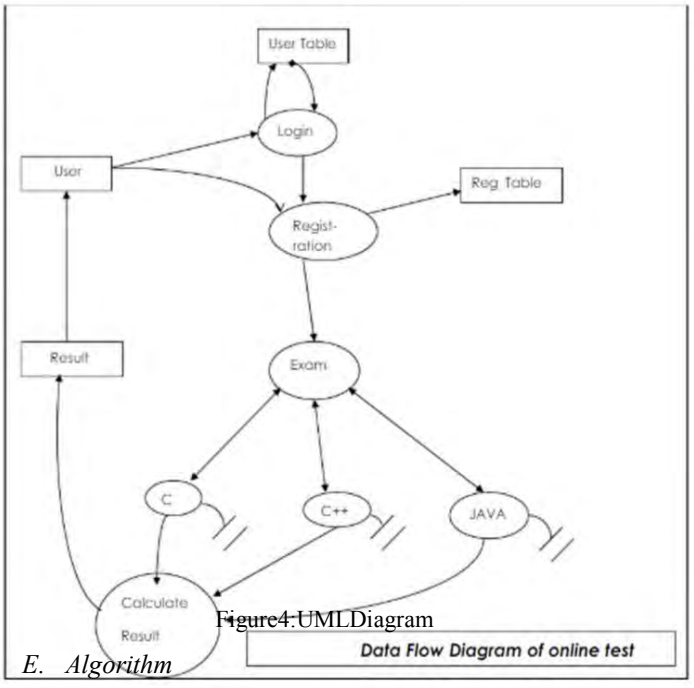


Figure4:UMLDiagram

E. Algorithm

- Step1:Student Identification:Theframeworkwill checkthepersonalityoftheunderstudybyutilizingenlistment beforeenteringthe testoftakenbyfacilitator.Thiswill likewisecheckqualificationoftheunderstudyfortest.

- Step2:AdminLogin:Theunderstudywillsigninto thespacewiththeclientnameandsecretphrasegavebythe administratorarealogin(Ex: username:mithil,Password: 123456).

- Step3:Speciallogininitotest:Theclientisrequiring composingtheclientnameandsecretphrase.Onthe off chancethattheclientidandsecretphrase isright,at that point theunderstudycangive thetest.

- Step4:GivetheExam: Theunderstudy willfinish the testthat is inthetestframeworkworkareawindow.

- Step5:Theadministratorneedstoembedthe inquiries.

- Step6:Random inquiriesandResults:Theinquiries arearbitrarilygiventotheunderstudies,understudypresents theresponsestotheserver;whentheclientsessionisfinished, the assessmentframeworkcreatestheconsequenceofthetest immediately.

- Step7:End.

F\ Flowchart

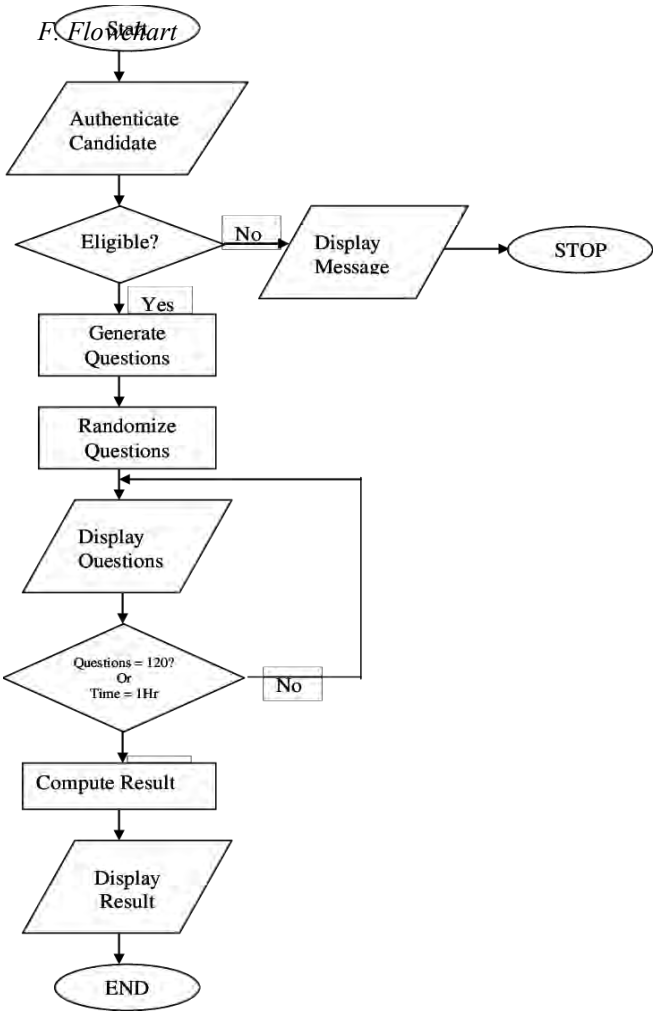


Figure5:Flowchart

III.RESULT

Figure6:LoginPage

IV.CONCLUSION

Itissignificantforassociationstoexploitthewebandrecei ve onlineassessmentframeworkfortheenrollmentof staff.This willexpandtheiroddsofgetting capablerepresentativesand furthermoretakeouttheobstructionof separationand inclinationinenrollmentofworkers.

V. ACKNOWLEDGMENT

We sincerely thank ourguideMr.Vikas Singh forhis guidance andconstantsupportandalsoforthesticktoour backs.Wealsothankstheprojectcoordinatorsforarranging thenecessaryfacilitiestocarry outtheprojectwork.

VI.REFERENCES

[1] Mohammad A Sarrayrih1, Mohammed Ilyas (2013) ,"Challenges of Online Exam, Performances and ProblemsforOnlineUniversityExamination".

[2] ZhaoQiao-fang,LiYong-fei(2012) ,"Researchand developmentof OnlineExaminationSystem"

[3] DeepankarVishwasKotwal,ShubhamRajendraBhadke, AishwaryaSanjay Gunjal,PuspenduBiswas(2016), "OnlineExaminationSystem"

[4] MohammadAttallahSarrayrih(2016),"Implementation andSecurityDevelopment OnlineExam,Performances andproblemsforOnlineUniversityExam"

[5] LinlinLi,SunLiangxu(2014),"OnlineExamination System for Microsoft Office Software Operations

Implementation Of Election System Using Block chain Technology

ShivaniKalewar ,Harsheen Kaur ,DivyaDubey, Mrs. Kalpana Gangwar

Computer Engineering

TCET,Mumbai, India

shivaani.1512@gmail.com,harsheenkaur23@gmail.com,diadubey14@gmail.com,kalpana.iiitm@gmail.com

***Abstract*— Blockchain is a singly LinkedList of blocks, with each block containing a number of transactions. It provides a decentralized, immutable data store that can be used across a network of users, create assets and act as a shared black book that records all transactions. Each transaction can be easily queried, affording greater transparency and trust to all parties involved. Blockchain technology has gained a lot of attention due to it's attractive features such as security, transparency and immutability in the area of elections and voting. Many systems Have been created for small applications like a boardroom voting process. BitCongress is an example of such an online voting platform based on the blockchain technology. It gives a whole system for voting based on tokens.Using this technology of blockchain, voting application can be created for maximum privacy and authentication.The election system in India is not completely transparent and secure. The details of the citizen are not securely handled and also control of the whole process lies in the hand of a handful of people resulting in a centralised system. This project proposes a distributed system on blockchain technology for secure and immutable voting for each user having an AadharID , i.e., a verified citizen of the country. It introduces a voting system which does not rely on a third party for tallying the votes and has a self- tally function.**

Keywords—Blockchain, Solidity, Ethereum

I. INTRODUCTION

Blockchain technology is the new age technology that can be used as a distributed ledger for storing data and also for making transactions. Elections and voting is an essential part of any democracy. With the development of technology, online voting systems have appeared and being actively developed. Electronic voting systems can be very convenient and can be accessed by all the citizens living in remote areas. Voting ensures that all the citizens are a part of the legislation of a country indirectly. Citizens have a say in who they want as a representative for themselves. Blockchain is a growing list of records called blocks which are linked to each other as a continuous chain. Blockchain is a decentralised system where the users are connected to each other and can also view the whole process. There are a set of miners who mine on the blockchain. These miners get an incentive for mining on the blockchain. Oncea transaction is made on the blockchain, to include it on the chain, it must be first verified by all the nodes on the blockchain. If maximum of the nodes verify it as a valid transaction, then only it can be added to the chain or it is rejected. A group of transactions are included into a block to add to the blockchain. Each block contains a cryptographic hash of the previous block, a timestamp and transaction data. This kind of process to add transaction to the blockchain ensures that each transaction is connected to the previous transaction. Once a transaction is done on the blockchain, it cannot be reversed ordeleted.

II.EXISTING VOTING SYSTEMPROBLEMS

Existing system for elections in India has a few problems which are addressed below :

- Vote counting is done by a third party which makes it a centralised system. This kind of system may arise in tampering with the finalcount.
- Votes given are not anonymous, i.e., voters can be linked to a particular vote. In order to maintain voter privacy, vote anonymity must beprovided.
- Proper authentication of voters or candidates during the election process is of utmost importance. If authentication is not provided, anyone can pretend to vote on behalf of a validvoter.
- Tampering with the Electronic Voting Machine (EVM) provided at the polling booth can happen by a particular party member or anyoneelse.
- One person must vote only once in a particular election. Multiple votes from a single person results in double voting issue.

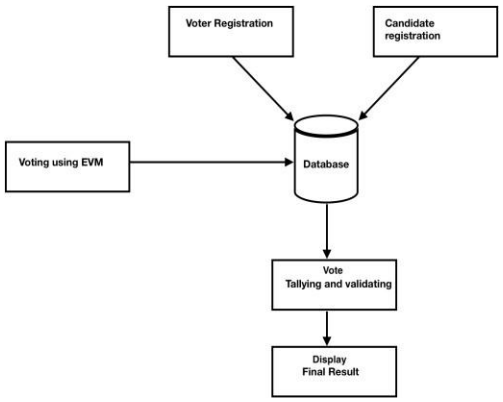


Fig: Existing system flow chart for voting

A) Objective:

The objective of this system is to create a secure and scalable environment for voting across the nation. It ensures that the election process of the country is as transparent as possible. The System must provide user authentication, it must provide verification and viewing of votes and it must prevent the double voting issues in current system. The major objectives of the proposed system are :

- Transparent and secure storage of details of all stakeholders of the system.
- Enable a secure anonymous voting and remove double voting.
- Transparent tallying of votes, which can be re - verified and talliedeasily.

III. PROPOSEDWORK

The proposed system introduces blockchain technology instead of using the regular database. Blockchain will ensure that once a transaction or a vote is cast, it cannot be tampered with and also the voters can be made anonymous. The voters and the candidates are all registered on the blockchain and the election officers verify the details of the users. Smart contracts written on the blockchain make sure that a voter can vote only once and also that the candidate has met all the conditions necessary to stand in the elections. The proposed system tries to overcome the shortcomings of the system followed right now in our country. It can be done by making use of the blockchain technology for complete transparency and anonymity and also makes sure that it is easy to use. There are four actors participating in this voting ecosystem. The following are the actors and their roles along with their functionalities in the system.

1. Voter Role: a voter will have the role of registering himself/herself on the blockchain. He/she has to provide necessary details for registering themselves on the blockchain. A voter also has the role of casting the vote on theblockchain.

2. Candidate Role: a candidate first has to register on the blockchain by providing the necessary details. Then to standintheelection,he/shemustapplyonthegiven election and provide his/her legitimate credentials. To apply for election, he/she must submit a security deposit. A candidate can also withdraw from the election. And also the candidate must register to a valid party.

3. Party Role: a party must first register itself on the blockchain. It must be a nationalised party and must have a party symbol. A party must also announce their representative candidate before the election process starts. A party can also dismiss any of its candidate if the candidate is notverified.

4. Officer Role: there will be two officers: election officer and presiding officer. Presiding officer will only create the election and watch over the election process. Election officer will be the one who verifies the voters, candidates and also the parties.

A) The proposed system can be divided into three kinds of modules :

1) User registration on the blockchain database:

There are three types of actors who will use the proposed voting system i.e., voters, candidates and parties. All the given three users must first register themselves on the blockchain database. All the relevant data regarding the users are to be submitted. Voters must submit their Voter Id for unique voting ID and also adhar card for verifying that the voter is a valid citizen of this country. Information like address, name, phone number, birth date, etc. must be asked by the system to store the relevant details of the voter. To ensure more security biometrics of the voter can also be taken so that the voter is fully authenticated. This information is stored on the blockchain and verified. There is also another actor called as the election officer who will verify the details of all the users of the system. The election officer will have access to the database. Candidates will have to register themselves as well. All the information that a voter has submitted must be given by the candidate as well as, a candidate must first be a valid voter. The candidate must also include the party he/she belongs to and also according to the Indian election system, a security deposit must be submitted before standing for any election. The deposit amount depends on the type of election he/she is standing in. This deposit fee must be taken at the time of registration of the candidate. Party is the third actor of this system. A party is a group to which a candidate belongs to. Each party must also register itself on the blockchain. Each party must have a symbol which a voter can recognise and vote for. Each party must mention the candidates belonging to theirparty.

2) The votingphase:

In this phase of the system, all the actors have registered themselves on the blockchain and the election can be started. There will be another officer called the presiding officer who will create the election and watch over the whole process. Once the election is created, the voters will be asked to vote for their choice of candidate. Before submitting the vote, the voter is first verified for authentication purpose. Once the verification is done, the vote is submitted for that candidate. A vote can only be submitted if that particular voter has not voted before andif the voted candidate is a valid candidate. If these conditions are not met, the vote is rejected by the blockchain. The whole election process can be seen by the candidates and parties for transparency.

3) The tally phase:

The last phase of this system will be the tally phase. This tally of votes has to be done along with the voting phase. If a third party is introduced for tallying, it can result in tampering of the final count if the third party is not honest. To avoid this kind of situation and to achieve full decentralisation, the tally is done along with the voting. When the vote is submitted, tally is also counted withthat.

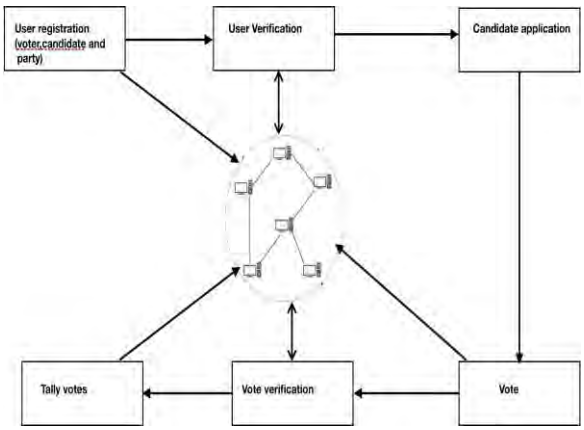


Fig: Overall workow diagram for proposed system.

IV. REQUIREMENTS

Software Requirements-

- Front End-HTML, CSS,JavaScript
- Back End-Solidity, SmartContract
- Operating System

Hardware Requirements-

- Min Memory:4GB
- Disk size: 1TB
- Processor: 64bit CPU with 8 cores and speedof 2.4 GHz

User requirements-
Officer-

- Creates and monitorselection
- Monitors voters andcandidates
- Depl oy result

Candidate-

- CreatesId
- Register himself/herself
- Gets vote count

Voter-

- CreatesId
- Castvote
- Observeresult

V. ARCHITECTURE

In the overall architecture of the online voting system, there will be four main entities. These four entities , i.e., Voter, Candidate, Party and Officer will be participating independently through the Blockchain. Each user will be linked to the Blockchain and all the transactions will be visible on the Blockchain.

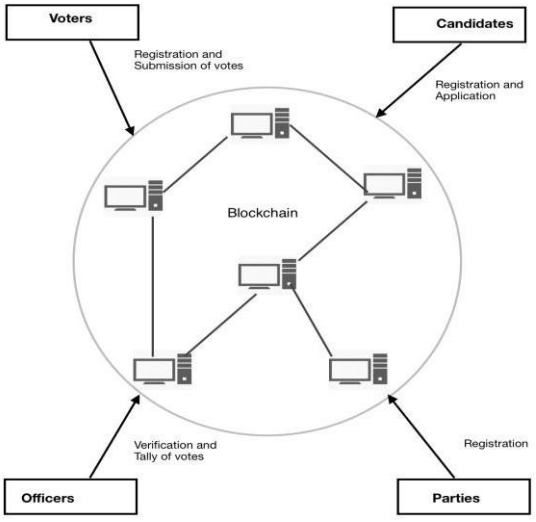


Fig: block diagram for proposedsystem.

VI. TECHNOLOGYUSED

1) Solidity:

Solidity is a contract-oriented programming language for writing smart contracts. It is used for implementing smart

contracts on various blockchain platforms. Solidity is a statically-typed programming language designed for developing smart contracts that run on the EVM. Solidity is compiled to bytecode that is executable on the EVM. With Solidity, developers are able to write applications that implement self-enforcing business logic embodied in smart contracts, leaving a non-repudiable and authoritative record oftransactions.

2) Truffle:

Using the Ethereum Virtual Machine (EVM) to create life simpler as a developer, a world-class development environment, testing framework and asset pipeline for blockchains. You get with Truffle: Built-in Smart Contract Compilation, Linking, Deployment and Binary Management, Automated Rapid Development Contract Testing, Scriptable, Extendable Deployment Migration Framework, Network Deployment Management for any amount of Public Private Networks, EthPM NPM Package Management, ERC190 Standard, Interactive Direct Contract Communication Console, Configurable build pipeline with tight integration assistance and external script runner runningscriptsinaTrufflesetting.Truffleisanenvironment developer, testing framework and blockchain asset pipeline. It enables developers to spin smart contract projectby clicking a button and offers you with a project framework, files and directories that make it much easier to deploy and test.

3) VSCode:

Visual code studio is a user friendly desktop application for compiling and running programs of various extensions. It has the solidity compiler as well.

3) MetaMask:

Metamask is a crypto-monetary wallet that can be used on browsers like Chrome, Firefox and Brave. It's also an extension to the browser. It operates like a bridge between ordinary browsers and the blockchain of Ethereum. The MetaMask wallet can be used on three distinct web browsers to store Ether and ERC20 tokens keys. It also enables users to browse a normal browser's Ethereumblockchain.

4) Ganache:

Ganache enables us to develop a personal Ethereumblockchain for testing, executing commands, and inspecting state while controlling how the chain works. It provides you the capacity to carry out without price all the activities you would have on the primary chain. Many developers use this to test their smart contracts during development. It offers easy instruments like sophisticated checks on mining and an integrated block explorer. Ganache offers 10 fake accounts and 100 ether for each account with a virtual network. Each account has an account address and also a private key associated with it.

VII. CONCLUSION

The current system used in India can be affected by double voting and intrusion of privacy of the voter. The proposed system provides privacy and transparency to the voters as the transactions can be viewed by all the users of the Blockchain. Data related to each and every election is safe and immutable as the data on Blockchain cannot be tampered by anyone and is irreversible. Blockchain provides advantages to all the stakeholders of the electoral system and also the citizens of the country. In a democracy like India, Blockchain technology can be effectively used. It can decrease the chances of double voting or false voting from the voter's side. Voters can validate their respective votes. Since Blockchain is transparent, it provides an unambiguous centralised system and also helps candidates to view the tally process themselves. It also helps reduce the time and energy spent by the officers for tallying as it is a self tally system and prevents the intrusion of duplicate voters. Blockchain helps create a transparent, secure and automated voting system for all the participants of the election process.

VIII. References

[1] SoumyajitChakraborty , Siddhartha Mukherjee, BhaswatiSadhukhan , KaziTanvi Yasmin Biometric voting system using adhar card in India , International Journal of Innovative Research in Computer and CommunicationEngineering.

[2] Gunjan Chaudhary, HimanshuGangwar, ShobhitBindal, Saurav Kumar Lakshya, Advanced SMS based voting machine.

[3] Patrick McCorry, Siamak F. Shahandashti, Feng Hao A Smart Contract for board- room voting with maximum voter privacy.

[4] Mrs.Harsha V. Patil, Mrs. Kanchan G. Rathi, Mrs. MalatiV.Tribhuwan,' A Study on Decentralized E-Voting System Using Blockchain Technology', IRJET Volume: 05 Issue: 11 | Nov2018

A Study of Live Streaming Practices in India

AdarshKumar Dubey, RahukKumar Jha, Salman Khan
Department of Computer Engineering
Thakur College of Engineering and Technology
University of Mumbai
Adarshdubey1998@gmail.com, Rj240499@gmail.com, iamjaankhan@gmail.com

Abstract - **Even though Live Streaming is on the rise in North America and Far East Asia, it's popularity in India is still negligible. Since 2016 there was an Internet Bloom in India because of the telecom revolution. As a result, there was a sudden upsurge in Live Streaming popularity. During the IPL Finals between Mumbai Indians and Chennai Super Kings in 2019, Hotstar a Live Streaming platform recorded 18.6 million concurrent viewers. To study this online gathering technological occurrence in India we conducted a survey of streamers and viewers. Based on their streaming, viewing and interacting practices we conducted an online survey of 500 people and even interviewed 10 streamers. The collected data revealed how different viewers react to different types of streaming content and streamers. We also observed the viewing perspective of viewers and the stimulus behind the streamers and even studied the interactions among the audience as well as between audience and streamers. Based on the collected data we are predicting the scope of live streaming in India.**

Keywords: live streaming, streamers, interactions, platform, reward-system.

I. INTRODUCTION

Live streaming refers to online streaming media simultaneously recorded and broadcast in real time. It is often referred to simply as streaming. Live stream services encompass a wide variety of topics, from social media to video games to professional sports. Platforms such as Facebook Live, Periscope, Kuaishou, and 17 include the streaming of scheduled promotions and celebrity events as well as streaming between users, as in videotelephony. User interaction via chat rooms forms a major component of live streaming. Platforms often include the ability to talk to the broadcaster or participate in conversations in chat.

Here we are creating a live streaming platform for gamers. We are using various tools and technologies like web technology, file transfer protocols like RTSP. This website will allow the streamer to stream their content and using this platform they can broadcast their content to users. It includes comment section used for comment, recommendation section, virtual money and many other features.

With live streaming becoming a financially viable market, particularly for esports, streamers and organizations representing them have looked for metrics to quantify the viewership of streams as to be able to determine pricing for advertisers. Metrics like maximum number of concurrent viewers, or number of subscribers do not readily account for how long a viewer may stay to watch a stream.

II. RELATED WORK

Twitch and Youtube gaming are two platforms where the gamers can broadcast their live matches with an audience which supports them through paid subscriptions and donations.

Twitch is available in US and requires subscription for $5 per month. Youtube gaming is a part of Youtube where normal users can watch their favorite gamers playing Esports. It has facility of Superchats and donations.

The highest ever live audience in Youtube Gaming was 400K when PewDiePie was playing Minecraft . Along with these 2, countries like China conduct Esports tournament which gather a large audience and generates revenue by sponsoring various companies.

III. TECHNOLOGY USED

HTML: HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects such as interactive forms may be embedded into the rendered page. HTML provides a means to create structured documents by denotingstructural semantics for text such as headings, paragraphs, lists, links, quotes and other items.

HTML can embed programs written in a scripting language such as JavaScript, which affects the behavior and content of web pages. Inclusion of CSS defines the look and layout of the content.

CSS: Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language like HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScriptBootstrap:Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front- end web development. It contains CSS- and (optionally) JavaScript-based design templates for typography, forms, buttons, navigation and other interface components React JS: -Reactis a JavaScript library for building user interfaces. React can be used as a base in the development of single-page or mobile applications, as it is optimal for fetching rapidly changing data that needs to be recorded. However, fetching data is only the beginning of what happens on a web page, which is why complex React applications usually require the use of additional libraries for state management, routing, and interaction with an API: Redux, React Router and axiosare examples of such libraries.

Node.js: Node.js is an open-source, cross-platform, JavaScript runtime environment that executes JavaScript code outside of a browser. Node.js lets developers use JavaScript to write command line tools and for server-side scripting—running scripts server-side to produce dynamic web page content before the page is sent to the user's web browser.

PostgreSQL: PostgreSQL, also known as Postgres, is a free and open-source relational database management system (RDBMS) emphasizing extensibility and technical standards compliance. It is designed to handle a range of workloads, from single machines to data warehouses or Web services with many concurrent users. It is the default database for macOS Server, and is also available for Linux, FreeBSD, OpenBSD, and Windows

IV. IMPLEMENTATION DETAIL

The project implementation is divided into 5 steps.

1. **Web Server with Basic authentication:**

We will place our react components inside the client directory and backend code in the server directory. For this part, we will be working inside the server directory. We are using passport.js for authentication. We have already installed passport and passport-local modules. Before we define our local strategy for authenticating users, let's create an app.js file and add the necessary code to run a basic web server. Mongo DB needs to be installed before doing this. We also need to define the schema for our User model. Create a database directory with UserSchema.js file in it and add the following code. We have three methods on our User schema. generateHash method will convert plain text password to bcrypt hash. We are using it in our passport strategy for converting plain password strings to bcrypt hash before storing them in the database. validPassword method will take in a plain text password and validate it by comparing it to bcrypt hash stored in our database. generateStreamKey method will generate a unique string that we will issue to users as their streaming key for RTMP clients.

2. **Setting up an RTMP server in Nodejs:**

Real-Time Messaging Protocol (RTMP) was designed for high-performance transmission of video, audio, and data between broadcaster and server. Twitch, Facebook, Youtube, and many other sites who offer live streaming accepts RTMP streams and transcodes then into HTTP streams (HLS format) before distributing them to their CDNs for high availability.

We are using node-media-server, a Node.js implementation of RTMP media server. It accepts RTMP streams and remux them to HLS/DASH using ffmpeg.

3. **Displaying live streams:**

For this part, we will be working in the client directory. Since its a react app, we will be using webpack and necessary loaders to transpile JSX into browser ready JavaScript.

We are using react-router for routing and bootstrap on the frontend along with video.js for displaying live streams. Add components directory with Root.js file in it and add the following code.

4. **Issuing streaming keys to broadcasters:**

Create Settings.js component.

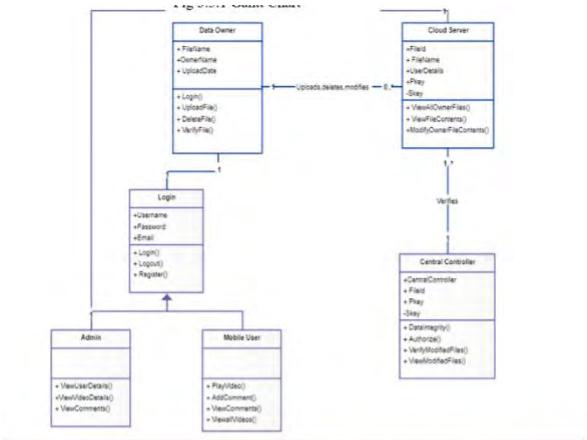
Inside our passport's local strategy, when a user successfully registers, we create a new user record with a unique streaming key. If a user visits /settings route, they will be able to view their existing key. When components mounts, we make an XHR call to the backend to retrieve user's existing streaming key and render it inside our <Settings/> component.

Users can generate a new key by clicking Generate a new key button which makes an XHR call to the backend to create a new key, save it to user collection and also return it so that it can be rendered inside the component. We need to define both GET and

POST /settings/stream_key routes. Create a server/routes/settings.js file and add the following code.

5. **Generating live stream thumbnails:**
In components, we are displaying thumbnail images for live streams. We will be generating these thumbnails whenever a new stream connects to our server. We will run a cron job to generate new thumbnails for live streams every 5 seconds. This cron job will execute every 5 seconds, retrieve active streams from NMS API and generate thumbnails for each stream using the streaming key. Import this job inside the app.js file and run it.

I. FLOWCHART AND RESULT



V. CONCLUSION

Live-Streaming platform will allow audience and gamers to connect in a very friendly and intriguing manner using the power of live broadcasting and live-chatting. It will revolutionize the gaming community like never before. It will help in smooth running of Esports tournament across the globe and will also allow advertisers to advertise in a very new way. It will also help to generate the revenue to the streamers by various means like ads, donations, live-streaming and sponsors.

References

[1] M. Papaloukas, "E-Sports Explosion: The Birth of ESports Law or merely a New Trend Driving Change in Traditional Sports Law?", Univ. of Peloponnese, May 2017

[2] Benjamin Burroughs and Paul Rama, "The eSports Trojan Horse: Twitch and Streaming Futures" , Journal of Virtual World Research, Oct 2015

[3] Marcus Shera, "The Social Effects of Twitch.tv" , George Mason University, May 2015.

Implementing Semantic Web For E-Commerce Website

Nikita Mahajan, Isha Jain, Rithvik Mandal, Siddhi Jain, Dr. Megharani Patil, Shivangi Dube
TE COMP A
nikitamahajan654@gmail.com , ishamadrecha@gmail.com , RITHVIKMANDAL2021@gmail.com , siddhirjain@gmail.com , megharani.patil@thakureduction.org , shivangidube77@gmail.com

Abstract—**Semantic Web Technologies enable machines to un- derstand and interpret data published in amachine-interpretable form on the web. In today's world where the smallest of the tasks require human-computer interaction, it is important that the data which is published on the internet along withbeing human understandable as well as machine understandable. This will improve the search results and will help the users in making informed decisions. Over 400 million searches are conducted every day on the Internet by people trying to find what theyneed. A majority of these searches include a web user looking forsomethingtobuy.Semanticsearchwillhaveadramaticimpact on the accuracy of these searches. It will possibly eliminate information asymmetry where a better informed buyer gets the best value. Currently, E-Commerce is critically vulnerable due to the scarcity of proper standards where HTML does not provide the syntax and semantics of information. Semantic web has the potential to influence the further development of the entire internet market, where e-commerce plays an important role. This paper discusses the semantic architecture for e-commerce using ontology language like RDF, in depth. It also discusses the shortcomings of the existing e-commerce websites and how semantic web will change the game for the internet market. This gives an overview of how the producer describes their resource online and how that data is made machine readable using the semantic web, which helps the consumer to retrieve efficientdata. There is a need for further developing these futuristic models based on our current understanding of e-commerce models and nascent semantic webtechnologies.**

Index Terms—semantics,ontology,RDF,e-commerce website

I. INTRODUCTION

A semantic web is a web that functions as an RDF knowl- edge graph. Google, Facebook etc. all have their versions of a knowledge graph. All databases contain entities and relationships. The entities are nouns (subjects and objects describing people and things) and the relationships are verbs, they describe how people, places and things are related toeach other. RDF graph databases address this shortfall by specializing in three things: 1) disambiguating the entities, 2) articulating the relationships between entities and 3) allowing large-scaleintegration.Thepointofthesemanticwebistomake the web more precise and more of a well-integrated whole that allows discovery at web scale. Ontologies, extensi- ble, flexible RDF graph data models are the main elements of semantic web that serve as rich description logic to live with the data and describe each individual context in a machine readable way. They make the data machine readable. Most companies use taxonomies, but some do use ontologies. Tax- onomies and ontologies both are hierarchical but ontologies goes further enabling any-to-any relationships, not just parents and children. Thus ontologies are more accurate and reflective of the natural world. Ontologies also allow inferencing, in which machines identify relationships between entities that are not spelled out in the initial data model, but which depend on relationships that were previously explicit. In this way, the whole graph can develop and grow at machine scale [1]. In today's world millions of people rely upon the information provided on the internet to make their purchasing decisions, but the ad hoc fashion in which the descriptions of products and services that are mentioned on the website are interpreted and offered up by mainstream search engines makes this process far less accurate than what the users might wish. As data moves out of the enterprise and onto the web the picture becomes quite less clear. Even on the product's own website the description and the information about the project is not conveyed properly. For the consumer wishing to compare and contrast video audio players from a number of competing providers, it is very difficult to ensure that they are comparing apples with apples. The web services aim to provide seamless interoperation among networked programs and devices. How- ever, the main obstacle of achieving interoperability among deployed Web services is that the computer is unable to interpret the data that is to be exchanged between the web services because the technical and functional description is based on semi formal free text language. The Semantic Web initiative's purpose is similar to that of the Web services :to make the Web machine processable rather than merely human processable. The main components of the Semantic Web are:a unified data model , • languages with well defined, formal semantics), and • ontologies of standardized terminology, used by semantically enriched service level descriptions The intent of this study is to enhance the usability and

usefulness of the Web and its interconnected resources by creating Semantic Web Services garden. The user needs to choose not only between brands but also between products and specifications and features. No person can have knowledge about all the different brands of products that are on the website. This may lead to the customer buying a product with unnecessary features. Knowledge graphs break down product features in a way that is easy for consumers to understand. The features and specifications of a product can be structured in such a way that is machine-readable. Users can then query this knowledge graph to find information and explanations tomake an informed purchasing decision. This simple example is one way in which we can implement the semantic web on an e-commerce site which will make it easier for customers to understandaproduct'sfeaturesandspecifications.Theinte rnet is made of millions and millions of web pages "stitched" together by links that humans (and also search engines) can follow. The search engines collect headings, meta tags and content, and create a way to help them gather and interpretthat data to say that your website is an e-commerce clothing store selling, specifically, clothes for men. This is where structured data comes into the picture and this is why it's important for everywebsitetohavestructureddataoneverypage.

II. LITERATUREREVIEW

E-Commerce framework is a strategic and excellent frame- work to bring your business online.E-commerce is a great and sustainable way to solve the problems as it contains the features which are not available in any canned solution available in market [5]. Semantics helps us to gather in- formation to solve this problem because with the help of semantics computer could understand the kinds of data it handles and how to interlink them in an intelligent way. This gathering of information will help us find better solution and structure to the problems faced. Network Security issue is a major part of information security framework which applied to components that affect to E-commerce security. The security threats or breaches or any kind of issue related to security is majorly due to guidelines for safe and secure online shopping through shopping web sites these are network related issues [5]. Using this data collected and processed through semantics we can learn how security issues can be solved in E-Commerce websites using different encryption techniques, public key, digital signature etc [5]. The semantic web vision, use different kinds of ontologies for different kinds of aims. The assessment of an E-commerce site for its semanticity requires a thorough analysis of its products or services. For example, to find a significant item over the net the customer can send few queries about the product and gather all the necessary information including its origin, make, description etc. The next step is to l analyze and look attheproductusingtheweb.However,inordertohaveamea ning, a product must have a relationship to other products, services etc. Therefore, there are many factors affiliated with the measurement of the semantic items. The main purpose of the semantics is to change the entire experience from general search to an intelligent search experience [6]. Hence toachieve this we can gather all the common information through the data collected. Each product will have a key attribute relatedto it.So with the help of similar key attributes and properanalysis of searches or any activities related to that data in the past we can give suggestions and make changes for the user experience respectively[4].

III. GAPIDEA

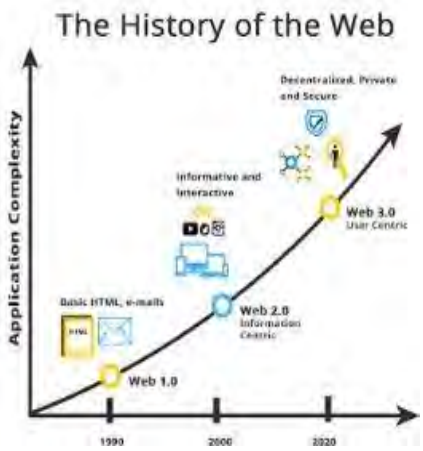


Fig. 1. **Timeline**

Semantic web is an attempt to intensify current web so that computers can process the information existing on WWW, understand and fix it, help humans to discover required knowl- edge. It is proposed to form a vast distributed knowledge based system to share data instead of documents.The aim of Semantic Web is to make Internet data machine-readable. To authorizetheencodingofsemanticswiththedata,technol ogies such as Resource Description Framework (RDF) and Web Ontology Language (OWL) are used.These standards promote regular data formats and exchange protocols on the Web, fundamentally the RDF.The Semantic Web is not a separate Web but it is an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation Advances in information technologies have strongly and consistently boosted organizations to deliver the correct information to the right person

at the right time. E-commerce websites are among the applications that have high demand of information sharing and reusing. The domain itself is ambiguous and complicated to humans unless its meaning is processed and supported by machines. Searching for a particular product is a multi-dimensional activity which synchronizes machine to machine, human to machine and human to humaninteraction.

Searching for quality Ethiopian commodities on the web will result into a unsuccessful query result. This is mainly because many products and services are unstructured and published in local languages like Amharic. Amharic corpora's, words and theories on the web are not modeled on the semantic web platform ontologically. Nowadays a lot businesses and companies are getting semantic web service enabled. Marketers of the future will need to understand their audience so well that they are able to create valuable experiences and information before the question is asked or the need is conveyed. Creating machine readable information content and inserting it into the existing commercial websites of the country will enhance the information retrieval facility by consumers on one side and will generate the opportunity to join these businesses to the linked data platform. Many researchers researched text mining and information retrieval approaches on Ethiopic or Amharic corpus. But almost all of them suggested the further development of ontologies in the respected domains.The semantic web annotation is open to each and every language. Amharic is the official language of Ethiopia, in which major public tenders are expressed, processed and evaluated with it. More than one hundred tenders are released every day, in which the majority of them are in Amharic language. Although it is believed that ecommerce is a bridge and a new way to an efficient global market, its simplicity, accessibility and flexibility needs a high demand for research. To sum up the motivation of this thesis is to represent this rigorous knowledge formally in a systematic manner and adapting it to the semantic web or the future web by making use of ontology language as a tool of gumption.

Currently, while conventional search engines are able to hunt for keywords on Web sites,only human users are able to read and interpret product and service information on the Web.The innovative technology used by the Semantic Web enables intelligent applications and search engines to grasp the meaningofthedata,processthisdataanddisplayitsothatitis comprehensible and meaningful to human beings. A practical way in which we can make use of the semantic web is to help consumers understand products and purchasing patterns.The simple answer is that the semantic web behaves as a sales assistant. Imagine going to buy a night camera on Amazon or any e-commerce website to monitor your garden.You need to decide not only between brands but also between products and specificationsandfeatures.Unlessyouhaveusedthesecameras before, you may not have a clue of what features to go for. Worse still, you may go for a product with many unnecessary features which are of no use to you. [1].With knowledge graphs, it is possible to break down product features in a way that is easy for customers to understand.The product data that is its features and specifications can be structured in such a way that is machine-readable. Users can then query this knowledge graph to find information and explanations to make their purchasing decisions. Possibly even through an automated chatbot.This simple example is one of the ways we can implement the semantic web on an e-commerce website to make it easier for customers to understand a product's features andspecifications.

IV. PROPOSEDSYSTEM

To measure information quality on semantic web ,we iden- tified the enhancements that semantic web can bring to e- commerce websites by studying the field of research. The semantic web provides a common framework that allows data to be shared and reused across application and enterprise.

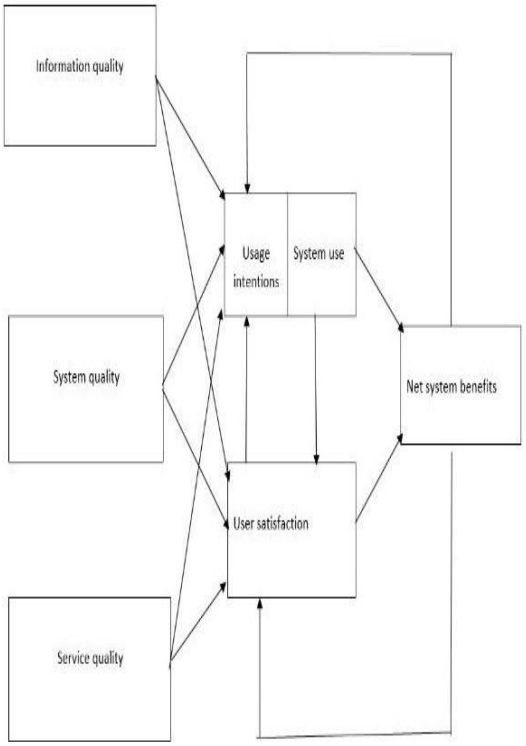


Fig. 2. **The representation of the model**

The machine-readable descriptions enable contentmanagers toaddmeaningtotheoriginalcontent,todescribethestruc

ture of the knowledge about that content. In this way, a machine can process knowledge itself, instead of text, using processes similar to human reasoning and inference, thereby obtaining results that are more meaningful and help the computers to perform automated information gathering and research .Good Relations improve search engine optimization for Google, Ya- hoo, Bing, and Yandex with schema.org, product information management, and e-commerce data quality management. The main question of our case study is whether the online con- sumers' satisfaction is associated with semantic enhancements oftheproducts'websitedescriptionintheonlineenviron ment.

V. METHODOLOGYANDTECHNOLOGY

The title Semantic Web is used to define a set of tech- nologies, tools and standards which form the basic construct a blocksofsystemthatcouldsupportthevisionofaWeb infused with meaning. The Semantic Web is a layered architecture, which is represented using a diagram which was proposed by Tim Berners-Lee, with many variations since. Figure gives a representation of thearchitecture.

Figure: Semantic Web layered architecture is a simplifica- tion which has to be used with some caution, it nevertheless gives a reasonable conceptualisation of the various parts of the

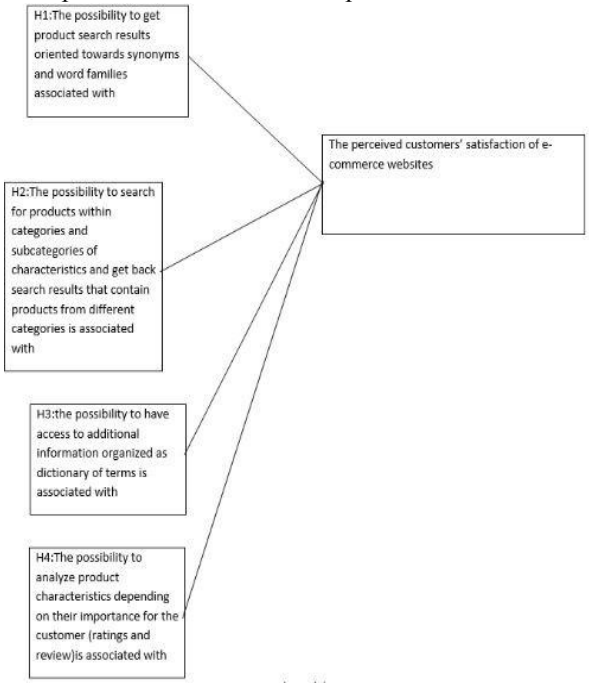
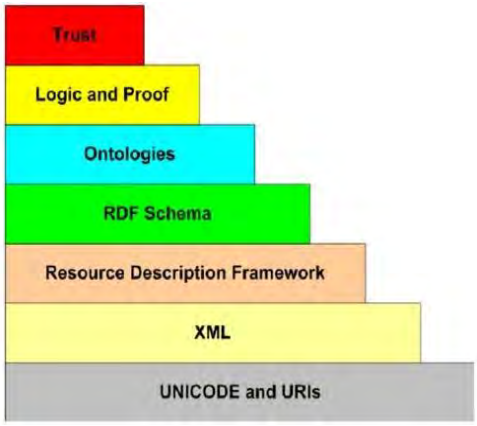


Fig. 3. **Research model**

Fig. 4. **Semantic web layered architecture**

Semantic Web. We describe briefly these layers: Unicode and URI: Unicode, is a standard for computer character represen- tation, and URIs, is the standard which is used for identifying and locating resources (such as pages on the Web), provide a base for representing characters used in most of the languages in the world, and for identifying resources. XML: XML andits related standards, such as Namespaces, and Schemas are very well established within the Web already but they don't com- municate the meaning of the data but are used for structuring the data on the web. Resource Description Framework: RDF isthefirstlayerofSemanticWeb.RDFisasimplemetadata representation framework, using URIs to identify and find Web-based resources and a 4 JISC Technology and Standards WatchSemanticWebTechnologiesgraphmodelfordescrib ing connections between resources. Several representationsthat are according to the syntax are available, including standard XML format. RDF Schema: A Simple modelling language usedfordescribingclassesofresourcesandpropertiesbetwe en them in the basic RDF model. It provides a simple reason- ing framework for deducing types of resources. Ontologies: A richer language for providing more complex constraints relations and complex knowledge about things, groups of things, and relations between things. Logic and Proof: A reasoning system working on top of the ontology structure to introduce new inferences or conclusions. Thus, using such a system, a software analyst or an expert can make deductions as to whether a particular resource satisfies its requirements (and vice versa). Trust: The final layer of the architecture addresses issues of trust that the Semantic Web can support. This component has yet not progressed far beyond a purpose of allowing people to ask questions of the trustworthiness of theinformationontheWeb,inordertoprovideanassuranceo f its truthfulness. Resource Description Framework (RDF), Web Ontology Language (OWL), and Extensible MarkupLanguage (XML), Web Ontology Language (OWL). HTML describes documents and the links that occur between them. RDF, OWL, and XML, by contrast, can describe arbitrary things, such as people, meetings, or any event. Business applications include improving information retrieval, thereby reducing information overdose and increasing the refinement and precision of the information retrieved. The machine-readable descriptions en- able content managers to include meaning to the content, to describe the structure of the knowledge about the content. In this way, a machine can process data and knowledge itself, instead of text, using processes similar to human

deductive reasoning and inference, thereby obtaining results that are more meaningful and helping computers to performautomated informationgatheringandresearchonitsown.



VI. CONCLUSION

Implementing the semantic web will change the game for the existing ecommerce websites. It will make it easier for the customers to search and compare various products as the descriptions will be machine understandable as well. This will helpthecustomersinmakinginformedpurchasingdecisions.

References

[1] A. F.Salam, A. F.Salam; Jason R.Stevens,IGIGlobal,Year:2007, Semantic Web Technologies and E-Business: Toward the Inte- grated Virtual Organization and Business Process Automation, ISBN: 1599041928,9781599041926,9781599041940

[2] https://www.woorank.com/en/blog/implement-semantic-web

[3] Semantic Web Technologies by BrianMatthews

[4] Building Ecommerce Framework for Online Shopping using Semantic WebandWeb3.0ISSN(Online)2278-1021ISSN(Print)23195940

[5] Gautam et al, international journal of advanced research in computersci- ence and software engineering "Network security issues in e-commerce ", March –2014.

[6]] Paul Kogut, Lockheed martin management data systems, Jeff Heflin- department of computer science and engineering "Semantic web tech- nologies foraerospace"-2003.

Transformation Of Rural India Through Digitization

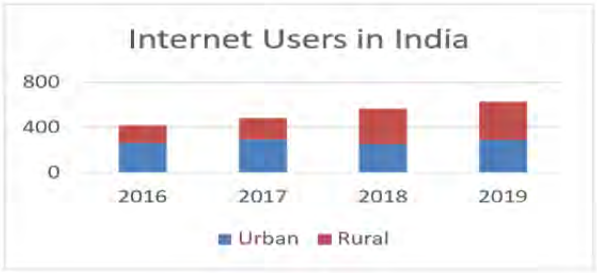
Sakshi Rai , Vij Patel , Prashant Upadhyay
B.E in Computer Engineering
Thakur college of Engineering and Technology
sakshirai0401@gmail.com , vijpatel22@gmail.com , prashant9upadhyay@gmail.com

Abstract:**The digital penetration within the Indian hinterland is growing silently but rapidly. The evolution and development in Information and Communication Technologies over the last decade has created a gap between urban areas and rural areas. Despite several initiatives being taken by the govt. NGO's, private social groups to use technology for efficient delivery of a spread of services, feasibility of such services is somewhere hindered by social or language barriers. So, in a developing economy like India, the success of any social development initiative relies on the government's involvement and active participation from both the private sector and people as well. India is continuing to create several success stories which will cause other initiatives by various interest groups.**

Keywords: Digital platform, Digitization, E-commerce, Rural markets, Digital literacy.

I. INTRODUCTION:

India's population now exceeds 1.3 billion with over two-thirds of its population and workforce residing in rural areas. Rural India contributes a considerable part of the total net value added in many sectors, with an overall 46% contribution to our national income. Studies conducted by IAMAI reflect some interesting patterns. 60% of those accessing the internet were using it for entertainment, 58% were using it for communications, 67% for online services, 65% for e-commerce and 49% for social networking. The number of internet users is estimated to cross 601 million internet users by June 2020 making India the world's second largest market for internet users. The purchasing aspirations of Rural Indians are often constrained by easy availability. Digitization and technology can facilitate access and availability of more and more services and products be made available to meet the rising aspirations of the underserved and unreached rural India. All these put together is now changing attitudes, awareness and lifestyles among rural Indians. A decade ago, the awareness, information and knowledge gap between the rural youth and his urban counterpart was significant which is now being narrowed significantly through the government's Digital India programme. One of the key factor supporting such development is the growing internet penetration, expected to grow from 25% in 2016 to 55% by 2025.



Internet Users in India

fig 1.1 Changing Face of the Indian Internet Users (2016-19)

II. BY IMPROVING DIGITAL LITERACY IN INDIA:

While, on one hand, the Govt is working towards creating the required infrastructure for e-enabling India, the other pre-requisite for realizing the vision of #Digital India is to make people digitally literate so that they can benefit from the action.

Illiteracy is a major obstacle to digital empowerment.
~ 'Offline and Falling Behind: Barriers to Internet Adoption' by McKinsey and Facebook
In villages, many children remain illiterate due to non-availability of schools in the vicinity, or due to

the lack of teachers in school.Illiterate people cannot derive any benefit from information or services available through text messages or other digital means, and so they depend on voice based programs for seeking and receiving advice.This holds true not only for people in the remote areas, but also for illiterate people in the urban areas too. My domestic help, for example, has little idea about the use of internet.

Conducting e-classes, and making
e- books or e-courses available in local languages can help in improving digital literacy, and gear up people for using internet to their advantage. Supporting the Digital India vision, companies like Intel and Microsoft will work the government to make 10 Lakh people digitally literate by the end of 2015.Intel India has launched the Digital Skills Training Application comprising of modules on Digital Literacy, Financial Inclusion, Healthcare and Cleanliness in five Indian languages. The company plans to impart digitalliteracy training to key resource persons in the first 1000 panchayats covered under NOFN rollout.

III. BY PROMOTING RURAL ENTREPRENEURSHIP THROUGH INTERNET KIOSKS:

Fig 2.Internet Kiosk in small towns | Image credit: dazeinfo.com

Providing training and capital to youths in rural/ semi urban areas to set up internet kiosksin the interiors will promote rural entrepreneurship.InternetKiosk with one or more computers, a tablet, Internet connection, with a web cam can be the set up in villages to be used as the hub of rural relatedness for providing education and training, information about agriculture and health care, employment news and market information. Gyandoot is an activity in Dhar district of Madhya Pradesh that connects villages throughout the district through numerous cyber-kiosks run by local entrepreneurs. More such activity in are needed in different districts throughout the country to promote the culture of rural entrepreneurship.

IV. BY IMPROVING TRANSPARENCY, EASE AND EFFICIENCY OF PROCESSES:

We know how tedious it used to be to apply for issuance or renewal of passport or any government related work we need an agent and it is time consuming and take lots of one's time and energy, before the process was e-enabled. Agents ruled the squat and there was no clarity on the process or the timeline. The online application and appointment for passport has streamlined the process to a great extent.E-enabling of processes such as filing tax, issue/renewal of driving licences, maintaining land records, results and shortlisting of candidates for employment schemes will improve the transparency, speed up these processes and save citizens from being bogged down by bureaucracy and corruption.

V BY ENABLING E-COMMERCE IN SMALL TOWNS AND RURAL AREAS

The small town people don't have more knowledge about limitation in the quality and variety of goods available in small towns or rural areas. Many products and brands are simply not available. The delivery of goods to consumer by bearer and postal services is also not very reliable in smaller cities, towns and rural areas. Setting up E-commerce hubs will enable people in small towns and farmers in rural areas to access markets, shop for better quality products, and get a better price for their produce without relying on middlemen. E-commerce can forward the selling of local goods such as herbal products, gardening supplies, traditional art etc while shopping for stuff such as electronics goods, solar energy panels etc.

VI. BY REDUCING RISK AND UNCERTAINTY:

Before setting out to sea, Madurai sits before a computer terminal collection information on the sea – like wave height, confusion and fish density. So do hundreds of other fishermen in Veerampattinam, a coastal village 15 km off Pondicherry city have the risk of collecting the fishes and also maintaining the cleaning process as well.

~ Wired to the future, India Today

The Information Village Research Project (IVRP) has redefined the lives of residents of villages in and around Pondicherry and also the surrounding coastal area of the sea.

Now, with the penetration of mobiles in rural areas, people can receive weather information or disasteralerts on their cells. This can especially be helpful in the coastal areas, where fishermen can be alarmed about weather warnings, or can be trained to check weather information on the net before dare to the sea.

VII. BY ENABLING DIRECT DISBURSAL OF MONEY TO BENEFICIARIES:

Financial involvement is another key area that can be impacted with the use of technology. Each year a vast number of people migrate from villages to cities to make a living. Any service that allows direct disbursal of money to beneficiaries will be immensely useful for people who want to remit money to their families living in thenative villages.A mobilewallet which allows consumers to send and transfer money, pay bills and do recharges, and take out cash, can greatly help in financial inclusion of people in the rural areas.

Vodafone's M-Pesa service for mobile money deportation, rolled out in April 2013, helps in disbursal of money of government schemes like the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) directly to the recipient in the circles of West Bengal, Kolkata, Bihar and Jharkhand.

Digitization can lead to a versatile development in India, by bringing the remote areas of the country within the fold of the mainstream economy. It is surely an ambitious plan, to realize the vision of IT +IT =IT; the benefits are aplenty, and so are the challenges. Nonetheless, with the hope that the challenges will be overcome, looking forward to the day when we have a virtually integrated #Digital India.

VIII. CONCLUSION:

The digital divide that exists between urban and rural India could become a thing of the past as we have experienced in the previous decades when it became feasible for every man to access television, cable and basic telephony through public booths. Thus the next decagon could catapult India to a new league of prosperity which will pave the way for elevating the living standards of sizeable number of rural poor and entitle them as never before if only we seize the convenience that digital mission provides.

References:

[i] Digital India – A programme to transform India into digital entrust society and knowledge economy, pib.nic.in/

[ii] An Overview of Digitization of Rural India and Its Impact on the Rural
Economy, LailaMemdani, abhinavjournal.com

[iii] Inclusion by mobile, Business Today

Phishing Detection Using Machine Learning a Review

Avinash Pandey,Anshika Pandey,Surbhi Pandey, Shiwani Gupta
Thakur College of Engineering & Technology,
Mumbai
pandeyavinash714@gmail.com,pandeysurbhi7@gmail.com,anshikapandey603@gmail.com,shiwanigupta3005@gmail.com

***Abstract*-- Phishing is a cybercrime, an identity theft used to lure users by posing as legitimate websites.Theseattackscanbethroughmails,messages or pop-up ads. Phishers target vulnerable users, particularly organizations, to extract sensitive information. Spear phishing is another form of phishingwhichisevenmoredangerousasitistargeted and carefully designed for a single recipient within an organization using social media stalking. As phishing detection grows more and more accurate, Phishers continue to develop even more legitimate looking replicas of websites. Even though there are manyanti- phishing applications or websites, Phishers still succeedatluringvictims.Phishingcanbesolvedusing, eitherBlacklistingapproachorbyHeuristicapproach. Blacklisting approach is a common solution but it is inefficient as Phishers continue to develop new fraudulent websites. In Heuristic method we extract various features of a URL to predict whether a particularwebsiteislegitimate,suspiciousorphishing. In this paper, we elaborate on Phishing attack and the approach to overcome this problem through URL feature extraction using MachineLearning.**

***Keywords*— Phishing, Heuristic, URL feature extraction, Machine Learning.**

I. INTRODUCTION

Phishing is a homophone of the word fishing which means catching a fish. There are four things here,afisherman,afishingrod,abaitandafish.The fishermanpatientlywaitsforthefishtogetattracted tothebaitwhichisattachedtotheendofthefishing rod.SimilarlyinphishingaPhisher(fisherman)uses a transmission medium (a fishing rod) to send a phishing message (bait) to the intended recipient or the victim(fish).

Phishing is a broader term, used to refer to a malicious attack intended to extract sensitiveinformation from individuals. Phishing is classified into many types, each having a different mode of transmission. Although Phishing sounds a very recentissue,ithasbeenpracticedsincetheinvention of computers. What makes it even more difficult to solve is that it relies on human vulnerability rather than technologicalvulnerability.Phishing detection on the other hand is a very difficultprocedureasPhishingiseverevolving.The reason for this, is that it is easy to copy an HTML code of any website, make slight changes in it and present it as a flawless replica of the original website. Phishing Detection can be done in two possible ways, one being creating public awareness butthenagainnotveryeffectiveandtheotheroneis through software. Software approach requires Data Mining technique, but it still is a complex process due to lack of dataset. The rationale behind this is that there aren't a specific number of features that characterize a phishing website. Hence shaping a dataset that covers all the features is difficult. Identifying phishing by common people is not easy as not every user is technically sound. Hence, Intelligenceagenciesarehiredbycompaniestokeep a watch for possible data snooping or snarfing. The most popular phishing detection is by the use of Machine Learning along with NeuralNetworks.

II. TYPES OF PHISHINGATTACKS

A. Vishing

Vishing refers to voice-phishing. Here the phisher impersonates themselves as a part of a trusted organization and tries to collect data over calls. Here a fake caller-ID can be used to give a more real appearance to it. E.g. A person acting asa bank employee asks you for your credit carddetails.

B. Smashing

Smashing refers to SMS-phishing or text phishing. Here text messages are used to send links of phishing websites. E.g. receiving offermessages from e-shoppingwebsites.

C. Spearphishing

One of the most dangerous form of phishing is Spear phishing. Phishing and Spear phishing are different because phishing is generalized while Spear phishing is targeted. Here extensive research iscarriedoutonthevictimandtheorganizationthey are a part of. It involves stalking, both physical and online.Thevictimisthensentaphishingmailortext based on their activity such that it becomes almost undetectable and unavoidable. E.g. A person uploadstheirlocationas"atStarbucks"andinstantly receives a message or email with the name of the same.

A .Whaling

WhalingisanotherformofSpearphishingwhich is used to target high-profile individuals. It is done to acquire high-level access to an organization which is only available by a company's board of directors. E.g. Targeting a CEO of acompany.

B. EmailSpoofing

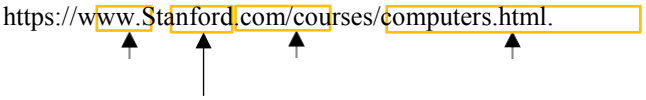
Email Spoofing occurs when an email header's from address is forged, so the recipient thinks that themailhasarrivedfromalegitimateemailaddress. Here the phisher can even impersonate a superior person or an organization. As a result, this is also called as Name Impersonation. E.g. receiving a mail from a superior.

C. MassTarget

Mass Target as the name suggests is intendedfor a large number of recipients who are sent an email or a message by the Phisher who impersonate themselves as a famous brand. The phisher collects information about a particular brand and then sends the customers their transaction invoices or giftcards etc. Due to this it is also called as Brand Impersonation. Here the URL usually contains spelling mistakes like amazon instead of amazon which isn't easily distinguishable.

D. URLManipulation

URL Manipulation refers to the manipulation of a certain part of the Uniform Resource Locator.The URL comprises of differentsections:



E.g. here https is the protocol, www is the subdomain, Stanford is the domain name and courses/computers.html is the path of the website. Since the domain name has to be registered, it can't be modified but the path and subdomain can be manipulated

With. Hence in URL phishing, the subdomainandpatharechangedtodirectthevictim to another website which looks almost the same as the originalwebsite.

III. PHISHING DETECTION

A. Blacklisting/Whitelisting

In this approach a predefined dataset is available which contains a list of phishing and legitimate websites. The drawback of this approach is that it cannot predict newly created phishing website. Asa result this method is highlyinefficient.

B. URL FeatureExtraction

This is a very efficient method of phishing prediction. This method extracts features which mainly characterize a phishing website and uses them to build a classification model. Data from known websites is collected and used to train this model. When any new website is accessed, it extracts its features and uses it to test the dataset to predict whether the given website is phishing or legitimate.

C. VisualSimilarity

1) Favicon: The phisher may copy the favicon (image associated with a particular website) anduse it in their phishingwebsite.

2) AddressBar:Thephishermayhidetheaddress bar with an image or script so that phishing URL is notvisible.

3) Appearance:ThephishermaycopytheHTML code of legitimate website to make a phishingone.

Visual Similarity based approach uses signatures to identify a phishing website. Signatures are basically features extracted from a whole legitimate website to compare it to a phishing one. A single signature may suffice to detect a phishing website.

D. ContentBased

Inthisapproachweexaminethewholecontentof a website to determine whether its phishing or not unlike the other approaches where we only looked superficially.Protocoldomainname pathsubdomain

TABLE 1

Comparison of Different Phishing Detection Approaches.

Approach	Advantages	Disadvantages
Blacklist/ Whitelist	Simple	Fails to detect newly created websites.
URL Feature Extraction	Highly efficient	Only takes into account the URL's and can be avoided. Frequently gives false positive results.

Visual Similarity	Effective image based anti-phishing	Only works if the websites are visually similar and doesn't take into account the code of the websites.
Content Based	Has the highest efficiency	Difficult to implement and time consuming.

IV. LITERARYREVIEW

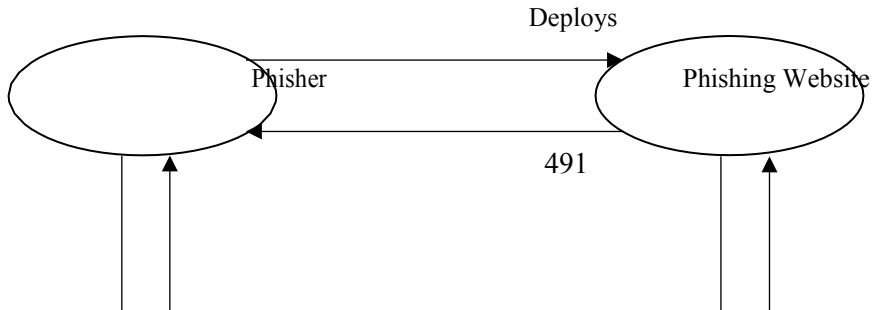
TABLE2

Listing and Gap Identification of Different Phishing Detection Methodologies and Approaches.

Reference Paper	Approach	Methodology	Conclusion
[1]	Content Based	Analysis of text-based content by itself.	TF-IDF approach: 97% of phishing sites and 6% false positives. Heuristics: 90% of phishing and 1% false positives.
[2]	Content Based	Novel heuristic based on URLs and website content.	Accuracy: 96.57% False Positive: 0.035%
[3]	Whitelist/ Blacklist	Creating a whitelist by registering the IP address of each site, which has Login user interface that the user has visited.	Classifies legitimate website as phishing when visited for the first time.
[4]	Whitelist/ Blacklist	Warns users on the web with a white-list of legitimate websites that is updated automatically	True Positive Rate: 86.02% False Negative Rate: 1.48%
[5]	URL Approach	Derivation of different components from the URL and computation of metric for each component.	Accuracy: 97%
[6]	URL Approach	Analysis of the lexical and host-based features of URLs	For Regression Trees: Success Rate: 91.08% (maximum) and Error Rate: 8.92% (minimum).
[7]	Machine Learning	Anti-phishing toolthat Can predict phishing attacks at a timescale using self-structuring neural networks(NN).	Accuracy: 92.48
[8]	Machine Learning	URL lexical Analysis and page content analysis.	Maximum Accuracy with ANN. Accuracy: 96.01% Approximate false positive rate: 0

PHISHING MECHANISM

In the fig. 1 given below, it is depicted how phishing mechanism takes place. The first step of phishing mechanism is the development of a phishing medium, in this case a website. In the case of spear phishing this comes after information gathering. This website is then deployed to the victim through any transmission medium available. The victim may or may not recognize that the website is malicious and may fall prey. The victim then enters their credentials into the portal. The data entered is then stored in the database available with the phisher. The phisher then uses this information to impersonate the victim and present themselves as the victim to the organization the victim is a part of. The phisher gathers information about the organization or uses this breach to implant any Trojan or virus in the company's entire workspace. Although this is the case only when the target is an organization rather than an individual. In the case of individual person, the information gathered is just used to perform unlawful transactions.



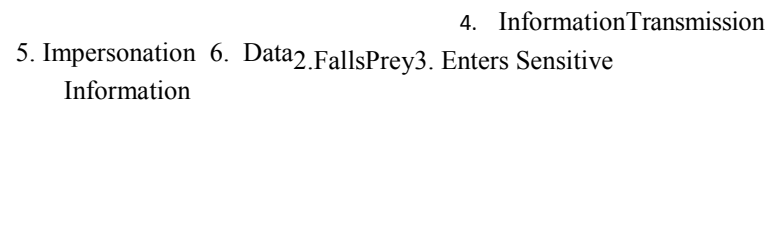


Fig. 1 Phishing Mechanism

V. URLFEATURES

TABLE 3

Major Features of a URL that can determine whether it is Phishing or not

Feature	Characteristic	Result
Length of URL	<54 characters	Legitimate
	>=54 characters <74 characters	Suspicious
	>=74 characters	Phishing
TinyURL	Present	Phishing
	Absent	Legitimate
HTTPS	Issuer Trusted	Legitimate
	Issuer not Trusted	Suspicious
	No Issuer	Phishing
IP Address in Domain Section	Present	Phishing
	Absent	Legitimate
@ Symbol	Present	Phishing
	Absent	Legitimate
Host name in URL	Present	Legitimate
	Absent	Phishing
DNS Record	Present	Legitimate
	Absent	Phishing
Domain Age	<6 months	Phishing
	>6 months	Legitimate
Domain Expiry	<1 year	Phishing
	>1 year	Legitimate

The above given table depicts the major features of a URL that can be used to depict whethera website is phishing or not. These features were identifiedbasedonthestudyofvariousphishingand legitimate websites. These features can be used to build a dataset to train and test a model. Although these features may not necessarily characterize all the phishing websites as there are many more features that can be manipulated with, these are the ones which can classify most of the phishing websites into the phishingcategory.

VI. PROPOSEDMETHODOLOGY

The following fig. 2 depicts a general methodology we propose to classify a website into phishingorlegitimate.Inthisprocesswefirstcollect a database of phishing and non-phishing website URLs to develop a whitelist and blacklist. We then extract the features of these websites. From these features we select the ones which are relevant to a phishingwebsite.Thenwepreprocessthesefeatures andconvertallthetextualdatatonumericform,tomake it easier to process them. This dataset is then split into train and test dataset and then given as an input to a classifier. This classifier then categorizes these URLs into phishing, suspicious and legitimate. This output is then evaluated using performance metrics. If the performance is low then hyper parameters are tuned and the process is repeated. Each of the following

processes is tried on various classifiers and the one with the optimum output is chosen. Then a new set of URLs are given to this final model to detect whether they are phishing or not.

A. Performance metrics

1) Jacquard Index (J): Calculates $(y,\hat{y})=\frac{|y\cap\hat{y}|}{|y\cup\hat{y}|}$

2) Precision: Calculates how many classes were correctly predicted.

similarity of actual values(y) and predicted values ($\hat{y}$).

$Precision=T.P.T.P+F.P.$

3) Recall: Calculates true positive rate. $Recall=T.P.T.P.+F.N.$

4) F-score: Weighted average of precision and recall.

$F_1= 2* \frac{Precision*recall}{Precision+recall}$

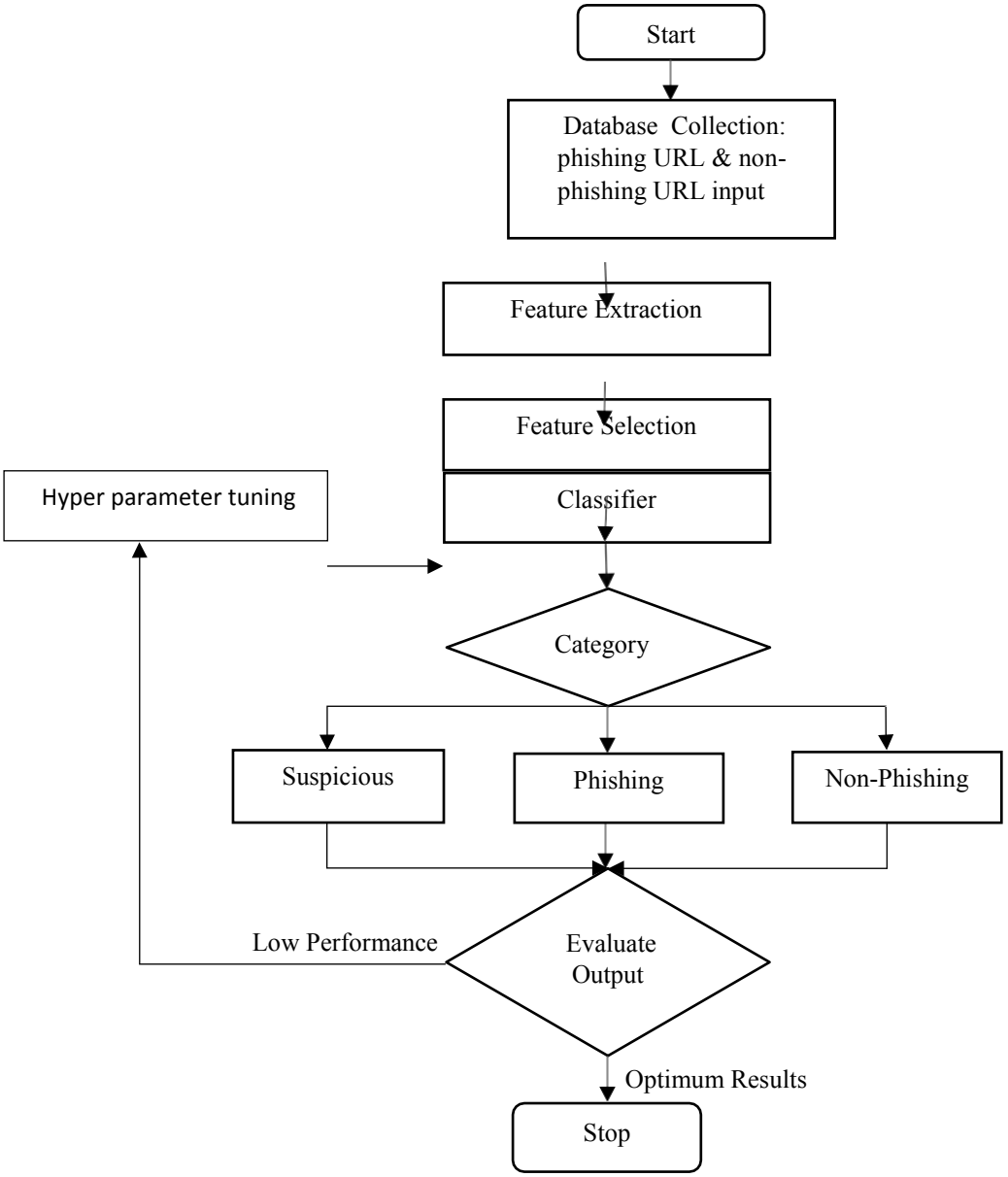


Fig. 2 Phishing Detection Mechanism

VII. Conclusion

In this paper, we first explained in detail about most of the possible types of phishing attacks that could

take place in today's world. We further gave details on how do they take place. We elaboratedon one of the most common method of phishing that is URL manipulation with an example. We also identified various types of phishing detection mechanism and their drawbacks. Then we focused on various URL features that are present in most phishing websites and can be used in phishing detection and their significance. We further proposed a general methodology that uses multiple classifiers and chooses the best one, to detect phishing. We listed various performance metrics to use to give better efficiency. This would result in higher phishing detection and lower security breach and data loss.

References

[1] Y.Zhang,J.I.Hong,andL.F.Cranor,*Cantina:AContent-based Approach to Detecting Phishing Web Sites*, New York, NY, USA, 2007, pp.639-648.

[2] R.S.RaoandS.T.Ali,*PhishShield:ADesktopApplication to Detect Phishing Webpages through Heuristic Approach*, ProcediaComputer.Science,vol.54,no.SupplementC,pp. 147-156,2015.

[3] YeCao,WeiliHan,andYueranLe.,*Anti-phishingbasedon automated individual white-list*, 4th ACM workshop on Digital identity management (DIM '08). Association for ComputingMachinery,NewYork,NY,USA,2008,51–60.

[4] Jain,A.K.,Gupta,B.B,*Anovelapproachtoprotectagainst phishingattacksatclientsideusingauto-updatedwhite-list*, EURASIP J. on Info. Security 2016, 9(2016).

[5] L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen, *Detecting phishing web sites: A heuristic URL-based approach*, in 2013 Inter-national Conference on Advanced Technologies for Communications (ATC 2013), 2013, pp.597-602.

[6] James, Joby & L., Sandhya & Thomas, Ciza. (2013). Detection of phishing URLs using machine learning techniques.

[7] R. M. Mohammad, F. Thabtah, and L. McCluskey, *Predicting phishing websites based on self-structuring neural network*, Neural Comput & Applic, vol. 25, no. 2, pp. 443-458, Aug.2014.

[8] R. Kiruthiga, D. Akila,, *Phishing Websites Detection Using Machine Learning*, International Journal of Recent Technology and Engineering (IJRTE), Chennai, Tamil Nadu, India, Sept. 11, 2019, 2277-3878 Volume-8, Issue-2S11.

[9] Joby James, Sandhya L, Ciza Thomas, *Detection of Phishing URLs using Machine Learning Techniques*, International Conference of Control Communication and Computing (ICCC), Trivandrum, Kerela, India,2013.

[10] A. MahaLakshmi, N. Swapna Goud, Dr. G. Vishnu Murthy, *A Survey on Phishing And It's Detection Techniques Based on Support Vector Method (SVM) and Software Defined Networking(SDN),* Hyderabad, India, Dec. 22, 2018, ISSN: 2249 – 8958, Volume-8,Issue-2S.

[11] Amani Alswailem, Bashayr Alabdullah, Norah Alrumayh, Dr.Aram Alsedrani,, *Detecting Phishing Websites Using Machine Learning*, Al-Imam Muhammad Ibn Saud Islamic University, Riyadh, Saudi Arabia, August, 2019, 978-1-7281-0108-8/19.

[12] anthana Lakshmi V ,Vijaya MS , *Efficient prediction of phishing websites using supervised learning algorithms*, International Conference on Communication Technology and System Design 2011, Elsevier Ltd.,Coimbatore- 641004, Tamil Nadu, India, 2011, 1877-7058.

[13] Jin-Lee Lee, Dong-Hyun Kim, Chang-Hoon, Lee , *Heuristic-based Approach for Phishing Site Detection Using URL Features*, Proc. of the Third International Conference on Advances in Computing, Electronics and Electrical Technology- CEET 2015, Republic of Korea, 2015, ISBN:978-1-63248-056-9.

[14] Waleed Ali, *Phishing Website Detection based on Supervised Machine Learning with Wrapper Features Selection*, (IJACSA) International Journal of Advanced Computer Science and Applications, Facultyof Computing and Information Technology, King Abdulaziz University Rabigh, Kingdom of Saudi Arabia, 2017, Vol. 8, No.9.

Web Security and its elements

Komal Singh,Aarya Singh,Tahera Shaikh
Department of Computer Engineering,
Thakur College of Engineering and Technology
singhkomal71199@gmail.com,Aryasingh1099@gmail.com,shaikh.tahera.i@gmail.com

Abstract:
Web Security is also known as "Cyber Security" involves protecting information by preventing, detecting and responding to web attacks
Today, technology is evolving at an unprecedented rate. These rapid advancements in technology make it strenuous to protect information at all times in an effective and efficient manner. The lack or absence of proactive security measures puts several individuals and businesses alike at risk of being subjected to cyber-attacks. It has been observed that the number of cyber-attacks is at an alarming ratein thelastfewyears.Thereisanurgentneedtoemploy effective cyber security measures, to prevent illegal and unauthorized acres and usage of our information systems. This includes protecting databases, computers and networks withspecialized and technical measures ofsecurity.

Keywords- Web Security, Web Threats, Web Prevention, Web securityalgorithms, Virus, Worms.

I. INTRODUCTION

Web application security is a branch of Information Security that deals specifically with security of websites, web applications and web services. Web sites are unfortunately open to security risks. And so are the networks to which web servers are connected. Web servers open a window between yournetworkandtheworld.Thecaretakenwithserver maintenance, web application updates and your web site coding will determine the size of that window, limit the kind of information that can pass through it and thus helps in establishing the degree of web security we can have.

II. HISTORY

A. 1970sthe first modern day hackers appeared as they attempted to mislead the system and make free phone calls, a practice that became known as "phreaking."
PerhapsthemostwellknowphreakedwasJohnDraper, A.k.a. Captain Crunch, who helped lay the practice.

B. 1980s

The 1980s saw the birth of computer clubs. This decade lead in the era of malware, marking the first virus, named "Brain", in 1986 as well as the ill-famed MorrisWormin1988.TheComputerFraudandAbuse Act was instituted in 1986.

C. 1990s

The 1990's brought the modern information security industry. Observable threats witnessed during this decade included the Michelangelo virus, Melissa, and Concept. Distributed denial of service attacks and thebotsthatmadethempossiblewerealsoborn,suchas Trin00, Tribal Flood network andStacheldracht.

D. 2000s

The first decade of the 21st century saw malicious Internet activity turn into a major criminal enterprise aimed at monetary gain. Malware also started and phishing attacks also became mainstream, first heavily targeting online banking then moving onto social networking sites.

III. WEBTHREAT

A web threat is any threat that uses the World Wide Webtoeasecybercrime.Webthreatsusemultipletypes of malware and fraud, HTTP protocols and other components, such as links in email or on servers that accesstheWeb.Webthreatsposeabroadrangeofrisks, including financial damages, identity theft, loss of confidential information/data, theft of network resources, damaged brand/personal reputation, and erosion of consumer confidence in e-commerce and online banking. It is a type of threat related to information technology (IT). Phishing is another commonthreattotheWebapplicationandgloballosses .

IV. TYPES OFTHREAT

A. Denial-of-serviceattacks denial-of-service attack (Do's attack) ordistributed denial-of-serviceattack(Dodosattack)isatrytomake a computer resource unavailable to its intended users.It

Generally consists of the concerted efforts to prevent an Internetsiteorservicefromfunctioningefficientlyorat all, temporarily orindefinitely.

B. Phishing

Phishing occurs when the attacker acts to be a trustworthy entity, either through email or web page. Victims are directed to fake web pages, which are dressed to look legal, via spoof emails, instant messenger/socialmediaorotherapproach.Oftentactics such as email spoofing are used to make emails appear to be from appropriate senders, or long complex subdomains hide the real websitehost.

C. Maliciousssoftware

A computer user can be confused or forced into downloading software onto a computer that is of malicious intent. Malicious software, is any software used to disturb computer operation, gather sensitive information, or gain access to private computer systems. Such software comes in many forms, such as viruses, Trojan horses, spyware, and worms. Ransom ware and Shareware is a type of malware.
BelowisagraphthatshowsthegrowthrateofMalware infection

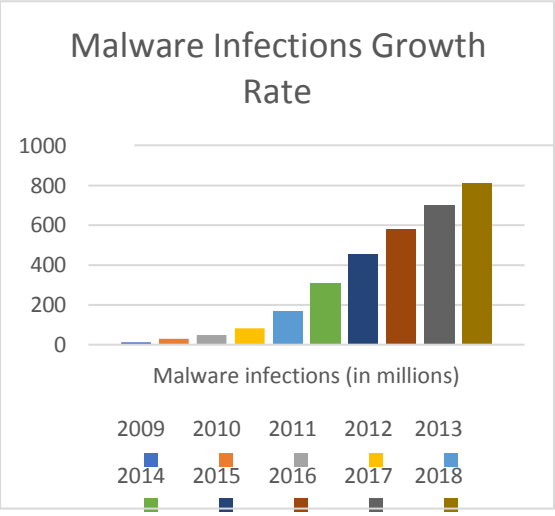


Fig. 1 Growth rate of Malware Infection

D. Application vulnerabilities

Applications used to access Internet resources may contain security vulnerabilities such as memory safety bugs or flawed authentication checks. The most severe of these bugs can provide network attackers full control over the computer. Most security applications and suites are incapable of adequate defense against these kinds of attacks.

V. PREVENTION

A. Network layersecurity

TCP/IP protocols may be secured with cryptographic methods and security protocols. These protocols include Secure Sockets Layer (SSL),Succeeded by Transport Layer Security (TLS) for web traffic, Pretty Good Privacy (PGP) for email, and IPsec for the network layer security.

B. Internet Protocol Security(IPsec)

IPsec is created to protect TCP/IP communication in a secure manner. It is a set of security extensions developedbytheInternetTaskForce(IETF).Itprovides security and authentication at the IP layer by transformingdatawiththehelpofencryption.Twomain types of transformation that form the basis of IPsec: the Authentication Header (AH) and ESP. These two protocols provide data integrity, data origin authentication, and anti-replay service. These protocols can be used alone or in combination to give the desired set of security services for the Internet Protocol (IP) layer.
The set of security services provided at the IPlayer includes access control, data origin integrity, protection againstreplays,andconfidentiality.Thealgorithmgives thesesetstoworkindependentlywithoutaffectingother parts of theimplementation.

B. Securitytoken

Someonlinesitesoffercustomerstheabilitytouse a six-digit code which randomly changes every 30–60 secondsonasecuritytoken.Thismeansthateverythirty secondsthereisonlyacertainarrayofnumberspossible which would be correct to validate access to the online account.

C. Electronic mailsecurity

Email messages are composed, delivered, and stored in a multiple step process, which starts with the message's composition. Then, the server opens up a connection(s) to the recipient mail server(s) and sends the message employing a process similar to that.

D. Message Authentication Code

A Message authentication code (MAC) is a cryptography method that make use of a secret key to encrypt a message. This method outputs a MAC value that can be decrypted by the receiver, using the same secret key used by the sender. The Message Authentication Code protects both a
Message's data integrity as well as its authenticity.

E. Firewalls

A computer firewall controls access between networks. It basically consists of gateways and filters which vary from one firewall to another. Firewalls also screen network traffic and are able to block traffic that is harmful. Firewalls puts restrictions on incoming and outgoingnetworkpacketstoandfromprivatenetworks. Incoming or outgoing traffic must pass through the firewall; only authorized traffic is allowed to pass through it. Types offirewall

- Packetfilter
- Stateful packetinspection
- Application-levelgateway

VI. WEB SECURITYALGORITHMS

A. AES (Advanced EncryptionSystem)

AESisasafeandreliablealgorithmwhichwhen employed, protects the confidentiality and integrity of data.AESisoneofthemostsecureencryptionstandards and thus, it is also one of the most widely used encryption algorithm. Breaking an AES algorithm necessitates computational powers that go far beyond the capabilities of the existing, moderncomputers.It is a symmetric cipher and thus uses the same secret code of 128 bits, 192 bits or 256 bits for the process of both encryption and decryption.AES is an iterative model and is based on substitution permutation.This algorithm utilizes a method for scrambling data. A key isusedtomixorscrambledatainacertainmannersuch that the new form of the data is unrecognizable. This form of the data can then be stored or transferred over a network securely. The jumbled data can only be unscrambled by individuals in possession of the key.AESemploysaseriesoflinkedoperations.Someof these operations replace their inputs by specific outputs.Each full chain or round, in the series of repeated ones, is made up of 4 primary sub-processes.
Process ofAES:

a. A fixed table is first referenced, to efficiently substitute the 16 input bytes. The thus obtained result comprises of a matrix of four rows and four columns.
b. A shift left operation is performed on each of the rows in the matrix. The first row is not shifted, the second row is shifted by one position, the thirdrow by two positions and finally, the fourth row is left shifted by threepositions.
c. Using a precise mathematical function, each column of four bytes is transformed. The four existing columns are take as input and four unrecognizable, new bytes are given out as the output. The resultant matrix consists of 16 new bytes
d. Thenew,calculated,16bytesofthematrixarenow considered to be of 128 bytes. This is then XORed to the 128 bits of the round key. If this is the final round, its output is considered to be the ciphertext. If not, the permutations are continued till the final round isreached.

B. RSA (Rivest, Shamir,Adleman)

RSA is an asymmetric cipher, thus it makes us of separate,distinctkeysforthepurposeforencryptionand decryption.Itisthemostwidelyrecruitedcryptosystem. Establishing a shared key is at times not possible or demands extra communications overhead. If oneperson hastocommunicatewithmultiplepeople,he/shewillbe forced to utilize multiple, distinct keys for each person. The sender would thus have to manage several keys and will have to send multiple messages to establishthem.

The use of an asymmetric cipher overcomes the above shortcomings and is a useful algorithm that makes use of cryptography at a large scale. It is based on simple yet clever concept.

RSA makes use of a public key as well as a privatekey, and the generation of these keys is the most complex part of its algorithm. Both public and private keys can beusedtoencryptamessage.Theprocessofdecryption of the message can be carried out by the opposite key, fromthekeyusedforitsencryption.Thereisanabsence of the need of historical use of the publickey.
Process of generating keys:

a. 2 large prime numbers p and q are selected. Using these, RSA modulus n is calculated using the following formula- n = p * q.
b. A derived number e is then selected. Here, e must be greater than 1 but less than(p-1)(q-1).
c. The pair (n, e) form the RSA public key. This key is then madepublic.
d. Then, the private key d is calculated using p, qand
e. For every n and q, there exists a unique d (private key). d is the inverse of e modulo (p-1)(q- 1).This security of this algorithm is acquired from the level od difficulty involved in factoring large integers, that are obtained from the multiplication of two large primenumbers.

C. Blowfish

Blowfish is a symmetric block cipher. Since its introduction, it has be conferred with inclusion in a huge number of encryption products and cipher suites.The encryption rate provide by blowfish cipheris impressive. It has also been observed that no other effective cryptoanalysis of it has been discovered to date. In terms of preference for usage, it rises above the other algorithms because of the fact that it has been placed in the public domain. It is non-proprietary and can be freely used by anyone.Blowfish has a 64 bit blocksize.Italsohasavariablesizedkeylength,which ranges from 32 bits upto 448 bits. This algorithm also makes use of 4 S-boxes or substitutionboxes.

D. Two Fish

The Two fish algorithm is derived from the Blowfish algorithm. This algorithm is considered to be a well-known algorithm employed for the encryption process. The scope of its usage lies within cryptrography and steganography. Twofish is used generally for the process of embedding. It can also be used for audio or images but has not yet claimed its efficiency in image steganography.This algorithm Has 128 bit blockcipher. A variable length key of upto 256 bits, is accepted by thiscipher.

VII. WEB SECURITYPRODUCTS

A. Antivirus

Antivirus software and Internet security programs can protect a programmable device from attack by

Detecting and removing viruses. Antivirus softwarewas mainly shareware in the early years of the Internet, but there are now several free security applications on the Internet to choose from for allplatforms.

B. Password managers

A password manager is a software application that helps a user store and organize passwords. Password managers usually store passwords encrypted, requiring the user to create a master password; a single, ideally very strong password which grants the user access to their entire password database.

C. Securitysuites

Socalledsecuritysuiteswerefirstofferedforsale in 2003 (McAfee) and contain a suite of firewalls, anti- virus, anti-spyware and more. They also offer theft protection, portable storage device safety check,private Internet browsing, cloud ant spam, a file shredder or make security-related decisions (answering popup windows) and several were free ofcharge.

VIII. CONCLUSION

History has taught us that we should never underestimate the amount of money, time, and effort someone will expend to hinder a security system.

The more secure we build the web today, the better and safer will be the future for all of us. Web services technology, is although very much in its developing stage, is all set to hit the mainstream enterprise computing. So, think before you click!

References

[1] http://www.ibm.com/support/knowledgecentr/SSPREK_ 7.0.0/com.ibm.isam.doc_70/amebW ebSec_guide_pdf.pdf

[2] https://en.wikipedia.org/wiki/Web_application _security

[3] https://en.wikipedia.org/wiki/Internet_securit

SmartTraffic Analysis-Using Time Series Analysis and Forecasting Methods for Big Data Analysis

PalakFuria, AnandKhandare
Department ofComputer Science,Department ofHumanities andSciences
ThakurCollege ofEngineering andTechnology(TCET)
palakfuria18@gmail.com,anand.khandare@thakureducation.org

Abstract-:
Intelligenttransportationsystem (ITS)isanadvanced technology whichenablesuserstobebetterinformed andmakesafer,more coordinated, and smarteruseof transportnetworkstomitigatetraffic congestion. This paperpresentsatrafficpredictionmodelbasedon time series.Laterusingdecomposefunction, random and seasonal traffic trendswereobserved.For thegiven datasetAIC techniquewasusedtochoose thebest predictionmodel. Using the statistical technique: ARIMAmoreaccurate andefficientresultswere achieved.Thus,theexperimentalresultdemonstratestha t theproposedmodelcanattain efficienttraffic prediction.

Keywords:- IT'S(IntelligentTransportation System),ARIMA (Autoregressiveintegratedmoving average), logistic regression,

I. INTRODUCTION

A. PROBLEMDEFINATION

Trafficestimationandpredictionsystems (TrEPS)have thepotential to improvetraffic conditions and reduce traveldelays.Tomonitor,manage,andcontrol the transportationsystem,these systemsusecurrently available and evolving computer,communication,and controltechnologies.Theyalsoprovidevariouslevels of traffic information and tripadvisory tosystem users, including manyITSservice providers, so thattravelers canmaketimelyandinformedtraveldecisions.Atraffi c prediction model that incorporates relevant results on viewingannual,monthly, and weeklytraffic trends.

Timeseriesanalysiscomprisesmethodsforanalyzi ngtimeseriesdata in ordertoextractmeaningfulstatistics andothercharacteristicsofthe data. Timeseries forecastingis the use ofa modelto

Predictfuture values based on previouslyobserved values. [1]

ARIMA isan acronymthatstands forAutoregressive Integrated Moving Average. An ARIMA model is a class of statistical models for analyzing and forecasting time series data. It clearly serves to a suite ofstandard structures in time series data, and as such provides a simple yet powerfulmethod formakingskilfultime seriesforecasts.

II. LITERATURE SURVEY

Becauseroadtrafficisthe visibleresultofthecomplex interplay betweentraffic demand(theamountof travelersmaking atripataparticularplaceandtime) andtrafficsupply (network infrastructure),when modelingit is usual to findthat the input–outputdata relationship isnoisyand thatthe relationshipsbetween thesevariablesaremultivariateand (highly)nonlinear (vanLintand vanHansberger,2012), additionally the processisusually highdimensional,non-stationary and tackledinreal-time

In the literature,there aretwo main approachesadopted for road trafficprediction:model-drivenanddata- driven. [2]

Model-drivenapproaches try to reproducetheroad network behaviorthrough simulation,anddepending onthe levelofdetailand the underlying traffic flow theory. (Barceló,2010;TreiberandKeating, 2013)

Ontheotherhand, thedata-drivenapproach aimsto reproduce the input–outputmapping butusually neglecting the underlying datagenerationprocess. (Ermagun and Levinson, 2016)

III. ARCHITECTURE

Thisprojectfocusesonfinding thebeststatistical-learningtimeseriesmodeltopredicttraffic.Itstartswith the collectionofdata.The **DataCollection**isa process bywhichtheresearchercollectstheinformationfromallthereleva ntsourcestofindanswers to the research problem.Thenextvery importantstepisdatacleaning. Data cleaningis the process of identifying and removing(orcorrecting)inaccuraterecordsfrom a dataset,table,ordatabase.Afterdatacollectionanddata cleaning, comes data analysis. Data analysis is the processofevaluating data using analyticalandlogical reasoning toexamineeachcomponentofthedata provided.Afteranalyzing data,predictionmodelsare createdusing AutoRegressionalgorithm.

3.1Algorithm:-forpredicting traffic trend logistic regression.Autoregressionisatimeseriesmodelthat usesobservationsfromprevious time steps asinputtoa regressionequationto predictthevalue at thenexttime step. Togetthemostaccurateresultvariousmodels wereused.Models suchas HOLTWINTERS, TSLM,

ARIMA,AUTOARIMAwereusedandtheb thebasis ofAICtest.
estandthe accurate resultswere choosingon

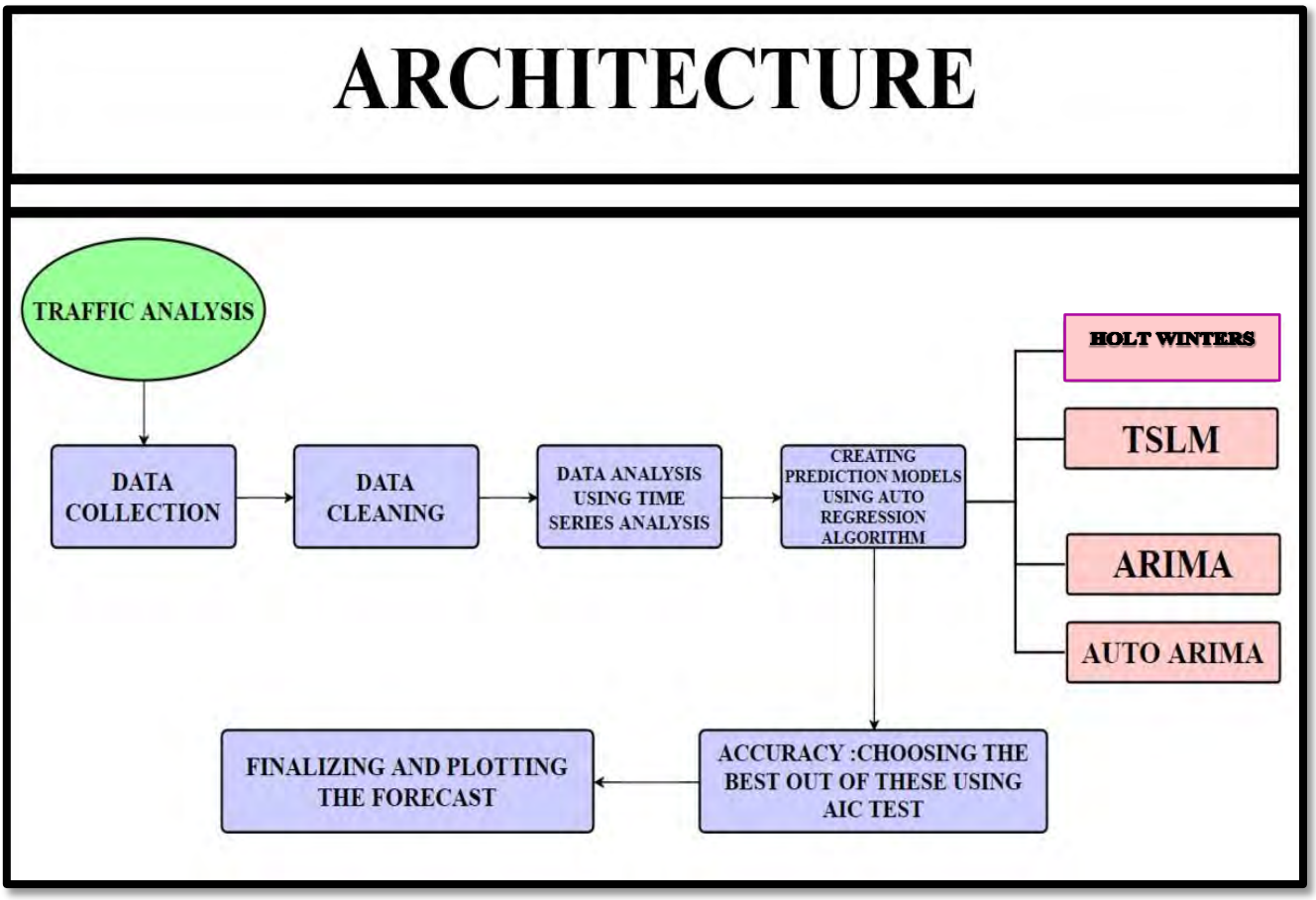


Fig1. Architecture

4.FurtherWork

4.1RecommendationSystem:-

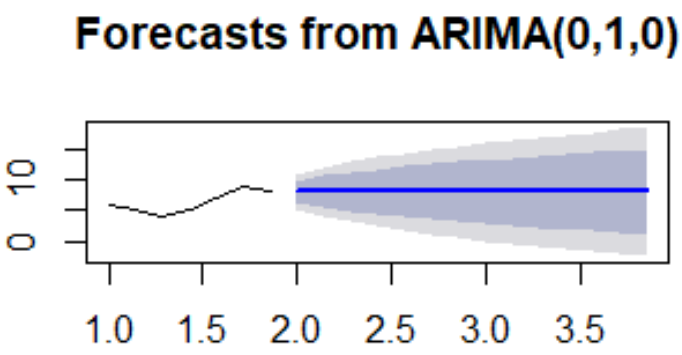
Toenhance the quality and tomakeituser friendly,a recommender systemis tobecreated wherethe userentersthemonth, dayoftheweekandthetime.Thesystemth en givesthe average trafficcount intherange of 1-10.

1-3 indicatesless traffic
4-7 indicatesmediumtraffic
8-10 indicatesheavytraffic

5.RESULTS

The end resultwasahigh-performancemodelthat producesananalyticalpanelprovidingtr afficprediction and analysisusinghistoricaltraffic.

Fig2.dailytrafficanalysis



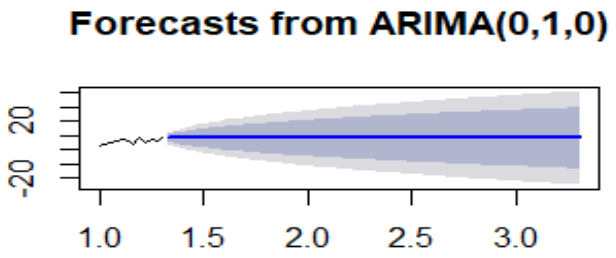


Fig3.weeklytrafficanalysis

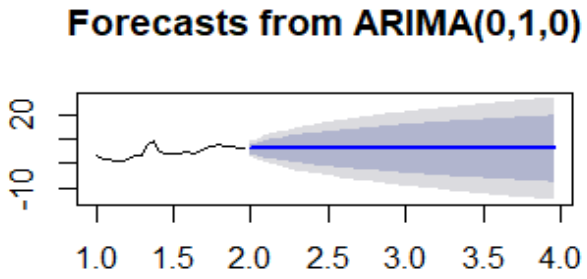


Fig4.monthly traffic analysis

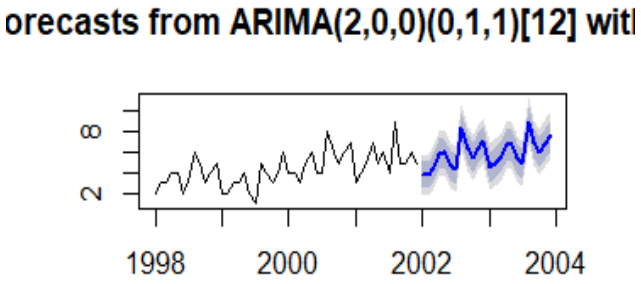


Fig5. Yearlytraffic analysis

Model	ME	RMSE	MAE	MPE
ARIMA	-5.670	55.213	45.611	-0.331
Box-Cox Transformation	213.268	235.805	213.268	10.388
Mean Forecast Method	866.998	868.128	866.998	42.255
Naive Forecast Method	2.814	44.356	37.086	0.090
Seasonal Naive Forecast Method	121.864	154.322	129.230	5.883
Neural Network	9.657	43.853	38.108	0.431

Fig6. Comparing results ofdifferentmethods

Aftertryingdifferentalgorithmsforthegivensetofdata wegotthefollowing accuracy resultforvarious methodswhichisshowninfig.6.

7. CONCLUSION

The trafficprediction iscriticalastraffic isenhancing dueto increaseuse ofvehicles. Theproposed literature usesARIMAmodelforprediction of trafficaccurately. Theresults arepredictedintermsofaccuracyandmean squareerror. TheaccuracyisenhancedsinceEuclidean distanceisused fordetermining theclosestdistance betweenthepointspresentwithinthedataset. The accuracy isobtainedby subtracting theactualvalue from theobtainedvalue.The leasterrorrateand enhancedaccuracy provestheworthofthestudy.In future,Geneticalgorithmcanbemergedwith the ARIMA forfurtherenhancementofaccuracy and reducingerrorrate.

8. References

[1]. NehaVerma HarjinderKaur-TRAFFIC ANALYSIS ANDPREDICTIONSYSTEM BYTHE USE OF MODIFIEDARIMA MODEL-Volume 8, No.8, September-October2017InternationalJournalof Advanced Research in ComputerScience

[2]. AttilaM.Nagy, Vilmos Simon-Surveyon traffic predictioninsmartcities-elsevier.com

Implementing System Forsign Languagerecognitionusing Machinelearning

Abhijeet Singh,Ankit Singh,Rashmi Singh,Nitin Harane
ComputerEngineering,TCET
University of Mumbai
ohhyessabhii99@gmail.com,ankii4050@gmail.com,ankii4050@gmail.com,nitinharane@thakureducation.org

***Abstract*—Handgesturesareusedasaway forpeopleto expressthoughtsandfeelingsit servestoreinforce informationdeliveredinourdaily conversation.Sign languageisa structuredformofhandgesturesinvolving visual motionsand signs, whichareusedasacommunication system. Forthedeafandspeech- impaired community,signlanguage servesasusefultoolsfordaily interaction.Signlanguageinvolvestheuseofdifferentparts ofbodynamelyfingers, hand,arm, head,bodyandfacial expressiontodeliverinformation.Thisposesa genuine communication barrierbetweenthedeafcommunityand therest ofthesociety,asaproblemyettobefullysolved untilthisday.**

***Keywords*— CNN(ConvolutionalNeuralNetwork),Hand motioncomparison,Signlanguage,Communication**.

I.INTRODUCTION

Handgestureisoneofthemethodusedinsignlanguagefor non-verbal communication.Itismostcommonlyused by deaf anddumb peoplewho have hearing orspeechproblemsto communicate among themselves or with normal people. Varioussign languagesystemshasbeendevelopedbymany makersaroundtheworldbuttheyareneitherflexiblenorcost. Effectiveforthe endusers.Patternrecognitionandgesture recognitionarethe developing fieldsofresearch.Being a significantpartinnonverbal communicationhandgesturesare playing key role in ourdaily life.Hand gesturerecognition systemprovides us aninnovative,Natural,Userfriendlywayof communicationwiththecomputerwhichismorefamiliartothe humanbeings.Byconsideringinmindthesimilarities ofhuman hand shapewithfourfingersandonethumb,Thesoftwareaims topresentarealtimesystemforrecognitionof handgestureon basisof detectionof some shapebasedfeatureslikeorientation, Centreofmass centroid,Fingerstatus,Thumb in position of raisedorfoldedfingers ofhand.

II.HOWITWORKS

Notebookdocuments(or "notebooks",alllowercase)are documentsproduced bytheJupiterNotebook App,which containbothcomputercode(e.g.python)andrichtextelem ents (paragraph,equations,figures,links.Notebookdocument s are bothhuman-readabledocumentscontaining theanalysis descriptionandtheresults(figures,tablesaswell asexecutable documentswhichcan beruntoperformdataanalysis.The JupiterNotebook Appisaserver-clientapplicationthatallows editingandrunningnotebookdocumentsviaa w e b b r o w s e r

TheJupiterNotebookAppcan beexecutedonalocal desktop requiring nointernetaccess(asdescribedinthisdocument)or canbeinstalledonaremote server andaccessedthroughthe internet.TheJupiterNotebook isanopen-sourceweb applicationthatallowsyoutocreate andsharedocumentsthat containlivecode, equations, visualizationsandnarrativetext. Usesinclude:datacleaningandtransformation, numerical simulation, statisticalmodeling, datavisualization, machine learning, and much more. JupiterNotebook is a web applicationthat allows youtocreateandsharedocuments.

A. CNN

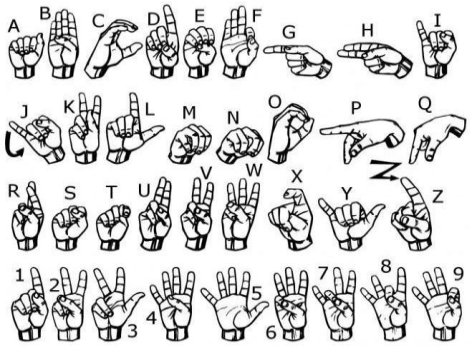
SignLanguageRecognitionusesneuralnetwork called3D ConvolutionalNeuralNetwork(CNN).The3DCNNmethodis based on2DCNNwhichisusedinspeechandimage recognition.The2DCNNimplementationis donebyextracting alayerto featuremap byusingkernelwindowcalled localreceptive field.Itreducesthe freevariablesandincrease generalizationcapabilityofthe network.The3DCNNis implementedusing2DCNNbyaddingthe motionfeature.The3D Convolutionisachievedbyconvolving3Dkernelwiththe cubeformedbystackingmultipleneighboringframestogether.12uses two3D-CNNintheirresearch,forextractinghandand

Upperbodyfeatures. EachCNNis3layersdeep.Theyalso applied localcontrastnormalization (LCN) forthefirsttwolayerandtheneuronsarerectifiedlinearunits(ReLU).2 1uses

2DConvolutional NeuralNetworkintheirresearch,with unusual featureintheirnetworkarchitecture.Theunusual features are ReLU Nonlinearity, multiple GPUs, LocalResponseNormalization, andOverlappingPooling.TheReLU Nonlinearityisthealternativeoftraditional neuronmodelsin CNN.Ittrainsseveral timesfasterthantraditional CNNandis bestfittobeusedinabigdataset.TheyalsousemultipleGPUs totraintheirdatabecauseofthesizeofdatathatdonotfitin oneGPU.

B. Dataset

Thedatasetformatispatternedtomatchcloselywiththeclassic MNIST.Eachtrainingandtestcaserepresentsalabel (0–25)as aone-to-onemapforeachalphabeticletterA-Z (andnocases for9=Jor25=Zbecauseofgesturemotions).Thetraining data (27,455cases)andtestdata(7172cases)areapproximatelyhalf the sizeof the standardMNISTbut otherwisesimilarwitha headerrow ofthelabel,pixel1,pixel2….pixel784 which representasingle28x28 pixelimagewithgrayscalevalues between0–255.



C. DataPreprocessing

Asthe datasethasalreadygivenCSVvaluesforimages,we don’tneedtodomuchpreprocessing.Ifdatasetoftheimage wasinrawformat, we have toconvertheminCSVformat arraysbefore doing anyofthefurtheroperations.Still,we performthefollowingsteps:

•Separatefeatures (784pixelcolumns)andoutput (resultlabel)
•Reshapethefeatures

D. Model

WewilluseKerastobuildthesimpleCNN(Convolutional NeuralNetwork).
Therearetotal7layersintheCNN:

1.1stConvolutional Layerwithrelu
2.1stMaxPooling
3.2nd ConvolutionalLayerwithrelu
4.2ndMaxPooling
5. Flattening
6. FirstFullLayerwithrely
7.OutputLayerwithsigmoid

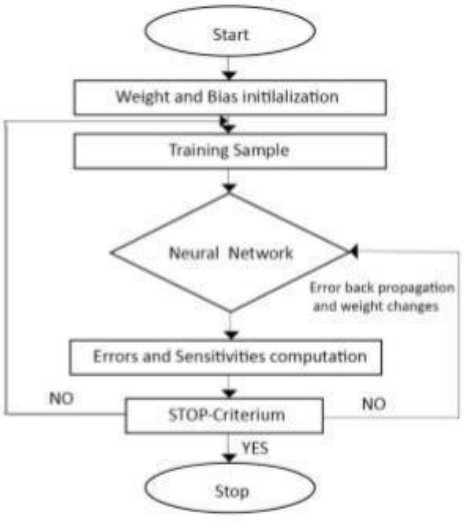
E. OpenCV

CreateaWindow:-
Wehavetocreateawindowtotakethe inputfromourwebcam.Theimagewhichwearetakin g asan inputshouldbe28x28 grayscaleimage.Becausewetrainedour modelon28x28 sizeimage.

F. Applications

Infuturework,proposedsystemcan be developedand implementedusing RaspberryPi.ImageProcessing partshould be improvedsothatSystemwouldbe able tocommunicate in bothdirectionsi.e.itshouldbecapable ofconverting normal languagetosignlanguageandviceversa.Wewill tryto recognize signswhichincludemotion. Moreover wewillfocus onconverting thesequence ofgesturesintotexti.e.wordand sentencesandthenconvertingitintothe speechwhichcanbe heard.

I. Flowchart



FigureI:-Implementationsteps carriedinthesystem

CONCLUSION

Signlanguagegesturerecognitionasit will behelpful todeaf people inassistingthemtointeractwithcomputers.However,it wasnoteasyformostoftheparticipantstoplacetheirhandon

Therectangular box,whichwasconsideredtobetheregionof recognition. Henceitisappropriatetorecommendthatabigger areaisrequiredsothattheparticipantswillnothavetostruggle toplacethe handintheregionbeforetheystartperforming the gestures.Forthesystemtobeeffectivelyappliedinacontrolled environmentlikeadeaf classroomthenawhitebackgroundcan beplacedbehindeveryseat of thecomputer user toreducethe challengeof complexbackgroundnoiseandimprovethe backgroundsubtraction.

ACKNOWLEDGMENT

Wesincerelythankour guideMr.NitinHaraneforhis guidance andconstantsupportandalsoforthesticktoourbacks.We alsothank the projectcoordinatorsforarrangingthe necessary facilities tocarryouttheprojectwork.

References

[1]Panwar.M. HandGesture Recognition System based on Shapeparameters.InProc.InternationalConference, Feb2012.
[2]Christopher Leeand YangshengXu. Online, interactivelearningofgestures forhumanrobotinterfaces. CarnegieMellonUniversity,TheRoboticsInstitute,Pittsburgh, Pennsylvania,USA,1996.
[3]Hyeon-KyuLeeandJinH.Kim.AnHMM-BasedThreshold ModelApproachforGestureRecognition.IEEE transactionsonpattern analysis and machine intelligence, Volume21,October1999.
[4]P.SubhaRajamandDr.G. Balakrishnan.RealTimeIndian SignLanguage RecognitionSystemtoaidDeafdumbPeople. ICCT,IEEE,2011.
[5]OlenaLomakina. Development of Effective Gesture RecognitionSystem.TCSET'2012,Lviv-Slavske,Ukraine, February21-24,2012.
[6]EtsukoUeda,YoshioMatsumoto,MasakazuImai,Tsulasa Ogasawara.HandPoseEstimationforVisionbased HumanInterface.InProc.10thIEEEInternationalWorkshop onRobot andHumanCommunication(Roman 2001),pp.473-478,2001
[7]Claudia Nölkerand HelgeRitter. Visual Recognitionof HandPostures.InProc.InternationalGesture

Workshop on Gesture-Based Communication inHuman

ComputerInteraction,pp. 61-72,1999.

[8]MeenakshiPanwarandPawanSinghMehra.HandGestu re

RecognitionforHuman ComputerInteraction.In

Proc. IEEE InternationalConference on Image Information

Processing,Waknaghat,India,November2011.

[9]Sanjay Meena. A Study ofHand Gesture Recognition

Technique.MasterThesis,DepartmentofElectronics and

CommunicationEnginnering,NationalInstituteofTechnology,

India,2011.

Predicting Player's Value in Ipl UsingData Science

Mayank Roongta,Kartik Rawool,Nisha Varghese
Department of Computer Engineering. Thakur College of Engineering and Technology
Mumbai, India
mayankroongta101@gmail.com, kartikr2k133@gmail.com, nishawishing@gmail.com

Abstract---Selectingateamisoneofthemostimportanttasksof any sport and cricket is not an exception to it. The best playing XI is selected for the match from available pool of the players by the captain. The captain identifies the different characteristics and statistics of each player while selecting the playing XI. Sometimes they are right and sometimes they go horribly wrong. This paper predicts the value of each player based on various algorithms and what each player brings tothe table for the team. The collected past performances of each player are evaluated and the form of the player is also considered while calculating each player'svalue.

Keywords---Data Science, Cricket, IPL, Standard Normal Distribution

I. INTRODUCTION

Cricket has been the most popular sport of India since independence and the emergence of IPL has taken this craze to the next level. In the IPL the domestic as well as the international stars are bought by the different franchises to playforthem.Sometimestheplayersturnsouttobebargains for the franchises but few times the players they havebought byshellinghugeamountsofmoneyturnsouttobehugeflops. This happens as the players are bought just on the basis of player's past reputation and ignoring the data such therecent performancesandabilitytoperformincertainconditions.B ut with the advent of data science some franchises have started using it and have made some really smart deals in the recent auctions. Hence in this paper we will predict all the player's value based on their past IPL performances as well as their recent form with the help of datascience.

II. REVIEW OF LITERATURE

Variousresearchpapersareavailableforthedataanalyti cs in cricket but most of them are for the team result prediction andveryfewarethereforthevalueofaplayerincricketand none of them for the player value in the IPL. Parker, Burns andNatarajan[1]definedamodelforvaluationofplayersfor IPL auction. Their model considered factors like previous biddingpriceoftheplayer,experienceoftheplayer,strikerat e etc. But it was way back in 2008 when IPL was season in nascent stage but now it has grown into the most popular cricket league in the world. The most similar research paper to our work is of Prakash, Patvardhan and Lakshmi[2]. They defined batting index and bowling index to rank players' performance for their models to predict outcomes of IPL matches. Bhattacharjee and Pahinkar[3] used the combined bowling rate(bowling average, strike rate and economy) to analyze the performance of bowlers inIPL.

III. PROPOSEDMETHOLOGY

We obtained all our data from popular cricket website www.cricinfo.com.Wecollecteddataofplayersfromt helast 5 years IPL for batting as well as bowling. The different parameters we collected in batting are the runs scored, averages, strike rate, etc. The parameters for bowling are wickets taken, strike rate, economy, etc. We imported all the data in MySQL tables to study on thedata.

After collecting the data for extracting the value of the player and also to bring all the parameters on the same scale thedatawasnormalized.Foragivenparameterwecalcu lated themeanandalsothestandarddeviation,andthennorma lized the value using mean and standard deviation. We appliedthe conditions that if the parameter is greater than the mean then we added 0.5 to it and if the normalized parameter less then meanthenwesubtractedtheparameterby0.5.Thiswasd one so that all parameters can be scaled and also so that we get where a player ranks in a group ofplayers.

Fortestingourcalculations,weusedadatasetofbats men, bowlersandall-rounders,eachdatasetincludedtop50players based on their runs and wickets. The dataset also includes importantparameterssuchasaverageruns,strikerate,n umber of matches played, number of innings, number of centuries and fifties, number of wickets, economy, etc. These parametersarecrucialinjudgingtheperformanceofapl ayer.
Also,thedatacollectedwasonlyregardingtheIndianPr emier League and did not include international

matches played by theplayer.

IV. DATAANALYSIS

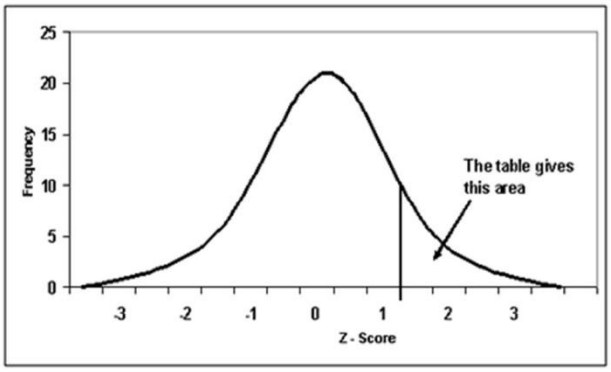
Brief introduction about the Analysis

Theanalysispartplayedthemostimportantinthispaper and hence it is better to get a brief introduction about it before startingtheanalysis.Sobasicallyforthispaper,wecollected the data and then started the analysis. The main parameters were average and strike rate for batsman and average, economyandstrikerateforbowlers.Soforunderstanding we haveshownacasestudywithdataofalltheplayersofoverall IPL seasons from 2008 to 2019. After importing the data, of top 50 players we calculated the mean and the standard deviation values of all the parameters. From mean and standarddeviationwecalculatedthestandardizednormalized value for each player with the help of standardized normal formula.

$Z = (x-m)/s$

x = the value that is being standardized
m = the mean of the distribution
s = standard deviation of the distribution

As the formula shows, a random variable is standardized by subtracting the mean of the distribution from the value being standardized, and then dividing this difference by the standard deviation of the distribution. Once standardized, a normally



distributed random variable has a mean of zero and a standard deviation of one. If we draw the graph with the standardized normal value it will look like this:

In this graph the z=0 at the peak or the mean value and z=1 at standard deviation value. This graph helps to find the probability of that normalized value. After calculating the z score of each player , a graph is hence drawn. A new value is calculated from the z score with the help of astandardized normal distribution table. Now here after calculating the value we consider 2cases:

Case 1: If z>0, we add 0.5 to the new value calculated from the table

Case 2: If z<0, we subtract the new value calculated from the table with 0.5

After getting the final values or probabilities, we convert it into percentile value by multiplying by 100. Suppose we get z = 0.4, we get 0.1554 from the table then we add 0.5 to it so it becomes 0.6554. On multiplying this probability by 100, we get 65.54 percentile value for z= 0.4. Hence we will give 6.554 to it in the range of 0 to 10.

A. Batsman

Whilecalculatingthevalueofthebatsmanweconsidered theformandexperienceofthebatsman.Themainparameters considered are average and strike rate. We devised an algorithm based on this and calculated Batting Power (BP1) of abatsman.

BP1=(0.4*((0.5*SR2019)+(0.5*Avg2019)))+
(0.25*((0.45*SR2018)+(0.55*Avg2018)))+
(0.15*((0.45*SR2017)+(0.55*Avg2017)))+
((0.1*((0.40*SR2016)+(0.60*Avg2016)))+
(0.1*((0.40*SR2015)+(0.60*Avg2015)))

In the above algorithm, the two main aspects of any player (form and consistency) are been taken care of by us. The performance of each player is calculated over 5 IPL seasons to consider each player's experience as well as consistency. Alsotheformistakenintoconsiderationbydecrementing the value of the performance over the 5 IPL seasons. The importanceofeachofthe5IPL(Total-1)isasfollows:

Performance in 2019 – 0.4

Performance in 2018 –0.25

Performance in 2017 –0.15

Performance in 2016 –0.1

Performance in 2015 –0.1

To calculate the performance of each batsman over the 5 IPL seasons, we have considered the 2 most important parameters in modern T20 cricket. They are:

1) Average

Average is considered to be one of the most important parametersinmoderncricket.Averageofabatsmanisdefined

asthenoofrunsbatsmanhasscoredineachinnings.Here for the above algorithm we have calculated the value of each batsmanusingNormalDistributionCurve.Wewillscaleeach Batsman value from 0 to 10. From the normal distribution curve,wegotthevalueofthepeak(meanvalue)inthecurve and set it as value5.

To help understand better, we have shown here the example of Average of top 50 batsmen in the entire IPL history.

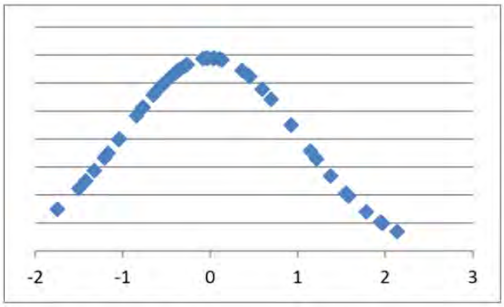


Figure 1 Normalized batting Average

As seen in the above example, the peak (mean value) is around30.Thataveragevaluewillbegivenvalue5basedonthepercentile.Thatmeaneachbatsmanhavingaverage30for that season will be rated 5 in thealgorithm.

2) Strike Rate(SR)

Along with average, strike rate is also considered an important factor. The strike rate is defined as the runs scored by a batsman per 100 balls. Similar to average, we have calculated the value of each batsman from 0 to 10.

The example graph for the strike rate:

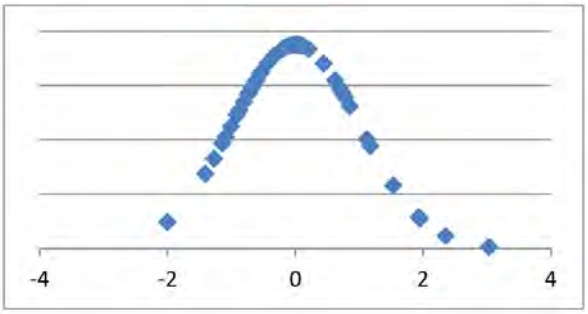


Figure 2 Normalized Strike Rate

Forthestrikerate,thepeakisatvaluearound130.Sothe batsman having strike rate 130 will be having value 5 in the above algorithm and accordingly the value for the remaining batsmen can be calculated with the help of standard normal distributioncurve.

3) Average VS Strike Rate

According to the various cricket analysts, over the years theimportanceofstrikeratew.r.taveragehasincreased.This hashappenedmajorlyduetomoreteamsoptingforexplosive players and giving less importance to the wicket of a player.
Soaccordinglythereisachangeinpercentagedistributionof average and strike rate over theyears.
Foryear2019–50%Averageand50%Strikerate
Foryear2019–55%Averageand45%Strikerate
Foryear2019–55%Averageand45%Strikerate
Foryear2019–60%Averageand40%Strikerate
Foryear2019–60%Averageand40%Strikeratepercentile we will subtract it by 10 to get the actual value of that bowler.

BP2=(0.4*((0.33*SR2019)+(0.33*Avg2019)+(0.33*Econo my2019)) +
(0.25*((0.33*SR2018)+(0.33*Avg2018)+(0.33*Economy20 18)) +
(0.15*((0.33*SR2017)+(0.33*Avg2017)+(0.33*Economy20 17)) +
(0.1*((0.33*SR2016)+(0.33*Avg2016)+(0.33*Economy201 6)) +
(0.1*((0.33*SR2015)+(0.33*Avg2015)+(0.33*Econo my201 5))

.

Intheabovealgorithm,thetwomainaspectsofanyplayer (form and consistency) are been taken care of by us. The performance of each player is calculated over 5 IPL seasons to consider each player's experience as well as consistency. Also,theformistakenintoconsiderationbydecrementingthe value of the performance over the 5 IPL seasons. The importanceofeachofthe5IPL(Total-1)isasfollows:

Performance in 2019 – 0.4

Performance in 2018 –0.25

Performance in 2017 –0.15

Performance in 2016 –0.1

Performance in 2015 –0.1

To calculate the performance of each bowler over the 5 IPLseasons,theparametersthatwehaveconsideredoverthe 5 IPL seasonsare:

1) Economy

Economy is defined in cricket as the no of runs conceded in 1 over (6 balls) by any bowler. Here

for the algorithm we will calculate the value for economy as per Normal Distribution Curve. From the normal distribution curve, we got the value of the peak (mean value) in the curve and set it as value 5.

B. Bowler

Whilecalculatingthevalueofthebatsmanweconsidered the form and experience of the bowler. The main parameters considered are economy and strike rate. We devised an algorithmbasedonthisandcalculatedBowlingPower(BP2) of a batsman. For bowlers while calculating value it will be oppositeofbatsmanastheminimumvaluesineachparameter meansbetterbowlersoaftercalculatingthevaluefromthe

The example Normal Distribution Curve for economy:curve, we got the value of the peak (mean value) in the curve

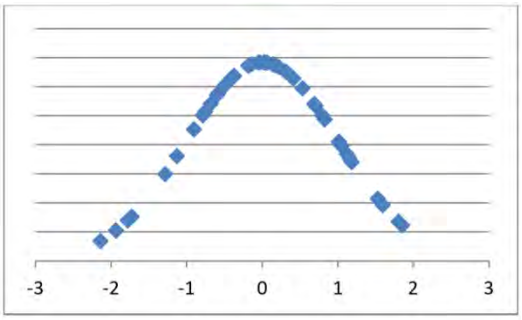


Figure 3 Normalized bowling Economy

Herethemeanvalueisaround7.8.Soitwillhavevalue5. But Economy of a bowler should be less for better value. Hence the calculating value of economy for each bowlerwill be(10-Value).Itmeansthatifabowlergetsvalue8fromthe curve,thefinalvalueforthebowlerwillbe(10-8)=2.

2) StrikeRate

Strike Rate of a bowler is the balls delivered per wicket takenbyabowler.Hereforthealgorithmwewillcalculatethe value for Strike rate as per Normal Distribution Curve.From the normal distribution curve, we got the value of the peak (mean value) in the curve and set it as value5.

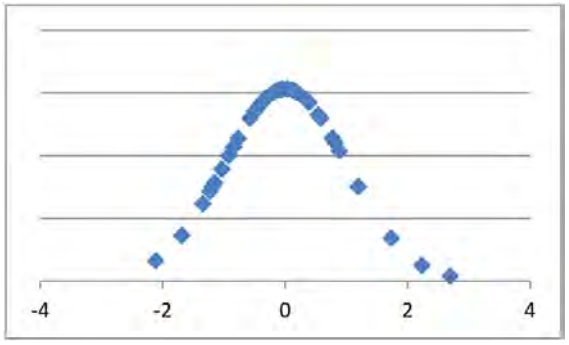


Figure 4 Normalized Bowling Strike Rate

The mean Strike Rate is around 21 and similarly standard deviationvalueisfound.Inthisaswellthevaluewillbe(10-valuecalculated).

3) Average

Average for a bowler in cricket is the number of runs conceded by a bowler per wicket taken. . Here for the algorithm we will calculate the value for Average as per Normal Distribution Curve. From the normal distributionThe example Normal Distribution Curve for Strike rate:

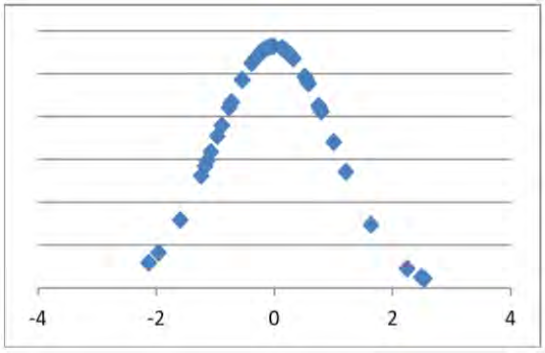


Figure 5 Normalized Bowling Average The mean value was found to be around 27.3.

4) Average VS Strike Rate VS Economy

According to the research paper of HemantaSaikia, DibyojyotiBhattajarchee, HoffieLemer [4] all the three parameters of each bowler should be given equalimportance hence0.33wasgiventoeachplayerinthecalculation.Butthe importance of each value has been fluctuating over time.But credible evidence has not been found to justify it. Hence the valueforeachparameterisconsideredtobethesameoverall the 5 IPLseasons.

C. ALLROUNDERS

InIPL,manyplayersdoeitherbattingorbowlingbut some of them perform both the skills on the cricket field. They are called as all rounders. There are

very few such players and hencetheyshouldhaveahighervaluethananysingleskilled player.Theallroundersarebasicallydividedinto3categories just for the convenience. Theyare:

1) *Batting All-Rounder*
These kind of all rounders have a batting value of more than 5 and bowling value of less than 5

2) *Bowling All-Rounder*
Thesekindofallroundershaveabattingvalueofless than 5 and bowling value of more than5

3) *ProperAll-Rounder*
ThesekindofplayersareveryrareinIPLbutthey are the most precious ones. They have a batting as well as bowling value of more than5

Algorithm to Calculate Value

Theyperformbothbattingaswellasbowling.Sobothe valueswillbeaddedheretoformAll-Roundervalue(AP).

AP = BP1+BP2

Calculation of Value (0-10)

So,understandingofvalueofeachplayerforthedifferent parameters (Average and Strike Rate for Batsman and Average, Strike rate and Economy for Bowler) is very essential to understand about the overall calculation. This value signifies about the percentile of players. For eg: If a batsmanhasavalue9.811calculatedaccordingtothegraphs, itmeansthatthebetteraveragethan98.11%playersthathave playedthatyear.Thevalueissimilartothebowlersaswell.

V. CONCLUSION AND FUTURE WORK

Selectingacorrectteamplaysaveryimportantroleinwinning acricketmatch.Hencethispaperplaysaveryhelpfulrole for that purpose as it provides each player's value irrespective of the player's past performances and reputation and helps to build a very strong team to win the IPL. In the above research paper, we have calculated each batsmen and bowler's value by a algorithm which consisted of different parts and the cumulative total of that algorithm gives the player's performance value. This is a unique kind of player thatcanhelpthefranchisestoselectsensiblyandsmartly.But therearevariousparametersthatarenotconsideredwhichcan enhance the algorithm. It includes Performance in winning matches, Performances in Knock-Out games and Fielding Performances. Also giving extra points to wicket-keepers, captains, all-rounders are the others aspects that can be discussed in the futurework.

VI. References

[1] D. Parker, P. Burns and H. Natarajan, "Player valuations in the Indian Premier League," Frontier Economics, vol. 116, October 2008.

[2]C.D.Prakash,C.PatvardhanandC.V.Lakshmi,"Data AnalyticsbasedDeepMayoPredictorforIPL9,"International Journal of Computer Applications, vol. 152, no. 6, pp. 6-10, October2016.

[3] KalpdrumPassi and Niravkumar Pandey," INCREASED PREDICTION ACCURACY IN THEGAME OF CRICKET USING MACHINE LEARNING", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.8, No.2, March2018

[4]HemantaSaikia,DibyojyotiBhattajarchee, HoffieLemer,"PredictingthePerformanceofBowlersinIPL: An Application of Artificial Neural Network", International Journal of Performance Analysis in Sports, April2012.

[5]HemantaSaikia,DibyojyotiBhattacharjee,AtanuBhattacharjee,"Performancebasedmarketvaluationofcricketers in IPL", Sports, Business and Mangement, May2013

[6] Siddhartha K. Rastogi , Satish Y. Deodhar," Player PricingandValuationofCricketingAttributes:Exploringthe IPL Twenty-Twenty Vision, April200

Interaction of Farmers with the Buyers

ChaitanyaKale,PratikGehlot Nitesh Mishra ,Chandan Mishra
Thakur College of Engineering and Technology Mumbai, India

***Abstract*- Farmers are the pillar of any prosperous nation. They contribute towards the major part of the economy but still they are deprived of the profit they deserve. India is a land where farmers contribute to 50% of the work force and GDP contribution of agro products is 17 to 18 % countries overall growth. The world is driven by technologies like Artificial Intelligence, Block chain, Data science, Web development, etc. and this list is endless. With such a huge leaps and advancement in modern technologies still farmers suffer the most in boosting their productivity of their crops and maximizing their income. They are still in the clutches of middleman, Rich Farm land holders, vendors who are not allowing them to get the maximum profit of their food grains and exploiting them .This paper analysis the various technological aspects and focusses on overall growth of farmers in terms of providing them opportunities to interact directly with buyers, vendors, restaurants, local market retailers with use of website, mobile apps. Not only farmers are deprived of their original cost but the buyers also don't receive these products at fair price which sometimes leads to burden on them.**

Keywords: [Farmers, technology, Profit, Smart farming, websites, Android apps, Machine Learning, Natural language processing, API]

I. INTRODUCTION

The overall growth of any nation takes place when there is equal share and contribution from all sectors of society. In 21st century technology is booming and connecting millions of people across the world. A smartphone devices have become such powerful that it can let your business grow from small market to a huge prosperous market. In terms of agricultural production activity of $375.61 billion, India stands on the second position for the largest producer of agriculture product. The total agriculture product output stands around 7.39 percent from India around the world. On the other part our neighboring nation china has contribution of $991 billion GDP in agriculture sector which is much more from India. Recent studies have revealed that fusion of agriculture and technology matter, and China is doing the right way. The use of artificial intelligence is one way that farmers are streamlining operations, increasing efficiency and ensuring sustainability."China has introduced a cloud-basedagricultural intelligence which has aimed to help the farmers of china. This will lead to increase of their crop yield and help them to achieve income of (71,289.40 Indian rupee.)".The need of such sustainable development is also needed in India where farmers are the prime contributor for the development. They are the ones who are grow food grain and feed the society but still they don't get enough cost of their crops. Every day around 35 farmers give up their lives which is very embarrassing for our nation where slogans are given such as 'JAI JAWAN JAI KISHAN'. We just need to make use of knowledge like machine learning, NLP, Deep Learning to make an application which has all integrated environment with all around development make use of these to grow the farm productivity. With the use of 'APP' farmers can connect themselves to the buyers and sell directly their products, Similarly with the use of 'websites' it can be feasible for the governments to launch their schemes and make people aware about development things they are intending for . This initiative taken by government will help them to reduce huge production of excess food grains which results in wastage sometime and also ensure the food security for the country and give farmer a feeling of self pride and satisfaction. Lowering the ceiling increases profit,and with AI advancements and collaboration in other areas, this goal is becoming a reality. The basic idea of using these systems are connecting farmers directly to the buyers allowing them to sell the quality product at reasonable price to the buyer and also ensuring that the maker get the price they deserve. Majorities of farmers are smallholders living in rural areas and thus lack amenities which will give them direct access to markets for their products and also they are deprived of agriculture market information. Therefore farmers are exploited by these intermediaries who offer low prices for agriculture products. Over the next few years, the application of these technologies will greatly shape the way the country manufactures food and distributes it to the most valuable sector of the nation. The country is already well on its way.

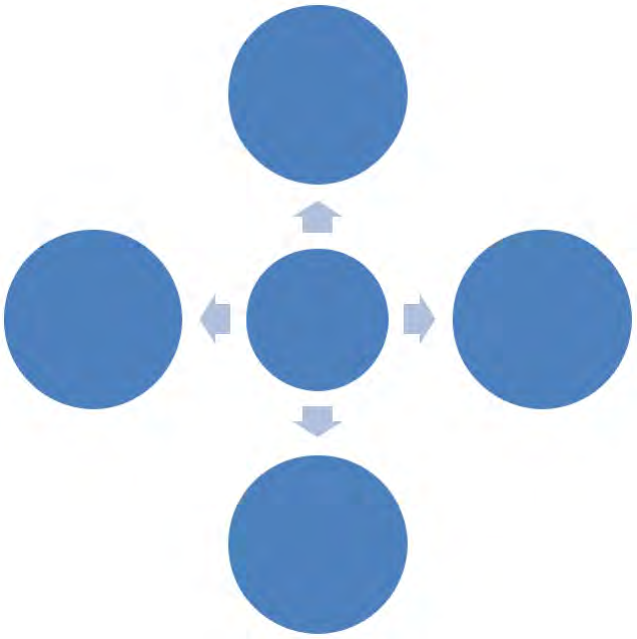
II. LITERATURE SURVEY

In India many attempts are made to improve the conditions of Indian farmers by the use of advance technologies. This would help the condition of farmers and farming. Studies are being carried to use implement E-farming with use of Artificial intelligence, cloud based technologies which will have a definite impact on the condition of farmers. It is proposed that there should be a web portal or android apps which would connect farmers to the buyer. These thingscannot be done individually by famers because farmers are not much educated so we also need special efforts from government and company professional in order to make famers aware of such existing platform.

III. IMPLEMENTATION

The system is formulated to improve farmers and buyer relation through an application based environment where famers can directly interact with buyers to sell their

products. In this farmers will grow their product and upload the details of their products so that buyer will come to know about the availability of the products. We make use of an integrated system where there will be facility of getting information about various schemes offered by the state and central government, improving the crop condition through the use of **Augmented Reality** by capturing the picture of **land** and make farmer aware what crops he should grow in repeated interval of time. In this application farmers will also get information whether he needs to grow crops in bulk amount or small quantity. This system also has an **AI** based **Chatbot** system which will interact with famer in his native language. Once the buyer will come to know about the product he can interact with famer through mobile Phones or **SMS** based **system** and make famer aware that he is interested in buying his product.



This system not only enables farmer to sell their products but also provides opportunities for vendors, restaurant owners, courier man, delivery agent to get employability through this. We also intend to have a SMS facility forthose farmers who cannot send their details through application.

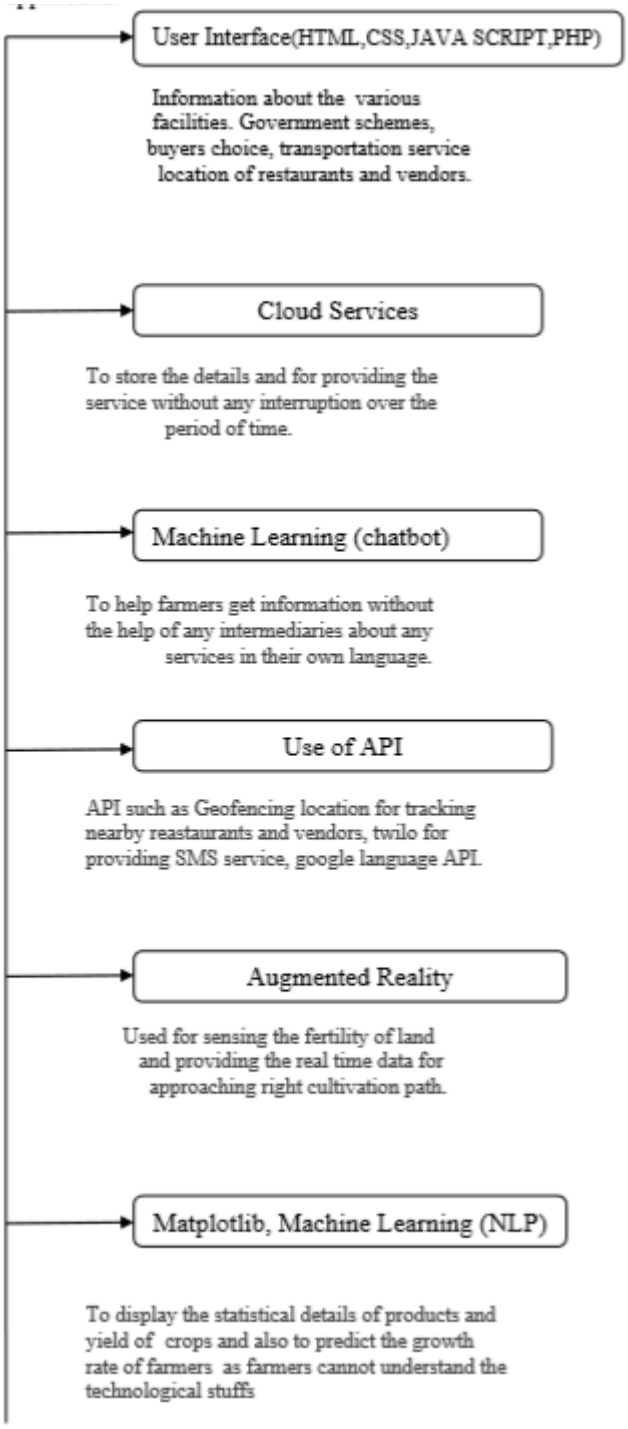
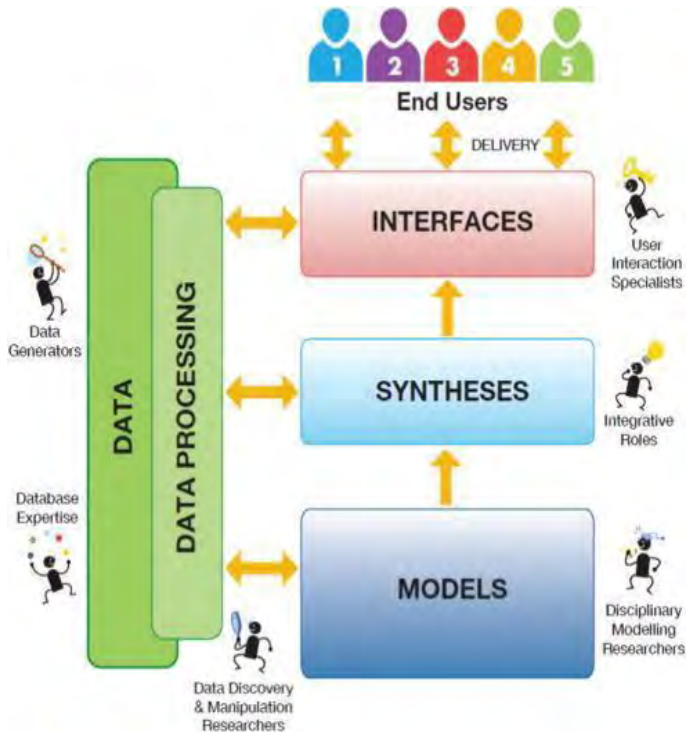


Fig A. Technology usage

IV. WORKING MODEL



We consider to have a model where there will be preprocessing of data so that some useful insights can be carried out, These insights can be used to distinguish important features by government so that they can bring some major reform in their policy.

V. USE OF API

The working of API is such that it allows to monitor the services based on ones requirement. We would be able to fence the nearby region around farmers where he can go and sell his product. Also we have API called Twilo which is used for sending SMS to the farmers. GOOGLE language API is used to translate language of web application in their own language.

TWILO API

```
// Your Account SID and Auth Token from twilio.com/console
$account_sid = 'ACXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
$auth_token = 'your_auth_token';
// In production, these should be environment variables. E.g.:
//     $auth_token = $_ENV["TWILIO_AUTH_TOKEN"]

// A Twilio number you own with SMS capabilities
$twilio_number = "+15017122661";

$client = new Client($account_sid, $auth_token);
$client->messages->create(
  // Where to send a text message (your cell phone?)
  '+15558675310',
  array(
    'from' => $twilio_number,
    'body' => 'I sent this message in under 10 minutes!'
  )
```

USE OF AUGMENTED REALITY FOR CROP PREDICTION

Linear Regression (LR) is used to establish relationship between explanatory variables (AR, AUC, FPI) and the crop yield as response variable. R^2 value clearly shows that yield is mainly dependent on AR. AUC and FPI are the other two factors influencing the crop yield. This research can be extended by considering other factors like Minimum Support Price (MSP), Cost Price Index (CPI), Wholesale Price Index (WPI) etc. and their relationship with crop yield.

```
import pandas as pd
d={'ARHAR':1 ,'COTTON':2 , 'GRAM':3 , 'GROUNDNUT':4,'MAIZE':5,'MOONG':6,
df=pd.read_csv('datasetFINALdata.csv')
from sklearn.model_selection import train_test_split
X=df[['Climate' , 'Year']]
y=df['srno']
X_train , x_test , y_train , y_test=train_test_split(X,y,test_size=0.3)
from sklearn.linear_model import LinearRegression
lg=LinearRegression()
lg.fit(X_train , y_train)
lg.score(x_test , y_test)
x_test
y_test
c=lg.predict(x_test)
```

This is the basic functionality what is to be included for implementing this model considering all the technological variations.

VI. AI POWERED CHATBOTS

An AI powered chatbots that can recognize the users need, collect different data from it and will be able to frequently asked questions and reply to it. Autopilots uses a task-driven programming model where tasks correspond to outcomes what users want through interaction with your bot. Some of the examples may be like making a call with buyer who has asked for his product, giving information about the market driven requirement products to be needed,etc. It uses natural language understanding(NLU) to detect what users are saying and match it to the appropriate task.

VII. USE CASES

There are several existing model which is trying to solve the problems of the farmers but it is segregated into different parts, overall there exists no model which tries to cover entire problem in one application. Our studies revolves around scenario where we want a model which has all the features involving from security, connecting government agencies, involving the concept of crop prediction, real time monitoring of the services and most important a level of satisfaction from the farmers and buyers.

Various different types of existing solution available in the market such as E-Nam, Kisan Network, AgriApp, KisanYojana,etc.

These all application services are inclined in offering one of the services where the need of time is a web application which has all features incorporated in it.

VIII. DISCUSSION

Currently this system cannot be implemented directly by the farmers as they have lack of access to technologies. We need support from Non government agencies and professional who will help us to setup the centre in village areas and organize awareness program to make people aware. There is an need of such system in the market as the gap between urban and rural areas is widening up. The recent survey shows that people are less involving in agriculture works because they are not getting access to markets and most of the products cost are eaten up by intermediaries.

Agriculture is one of the activity which is done for feeding the large humanity but still if the makers are in trouble it is one of the saddest part of our society. Every year there are many farmers who give up their lives due to loss of crops, huge storage of produces also lead to wastages some time.

Farmers are not only the ones who suffer but the buyers too have to bear hefty cost of the vegetables, grains, etc. There is always a mismatch between the original price laid by the farmer and what a buyer pays. This system does not justifies to the equal share to be given to farmers and buyers.

IX. FUTURE SCOPE

This system can be used a model which can be used to form an organization which will maintain all the ecosystem and ensure that this application is maintained and updated with growing technologies. We also would like to include advance security system based on Blockchain which can be use to link payment option with the buyers. There are also multiple things which can be included along with this model.

X. CONCLUSION

The study evolves around how can we help farmers with modern technologies and making them independent in terms of growing and selling. We want to incorporate technologies such as Artificial Intelligence for chatbots, NLP for deducting references from regression model, having a web application with features such as providing interface for getting information on different things. We would like to have collective support from all the working bodies I bringing up this model. This is not an individual task but as a whole society we need to implement this.

References

1. Transdisciplinary Approach for Sustainable Rural Development by RajkumarMurugesan , S . K. Sudarsanam

2. Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology SanderJ.C.Janssen[a]CherylH.Porter[b]Andrew D.Moore[c]IoannisN.Athanasiadis[a]IanFoster[d]JamesW.Jones[b]JohnM.Antle[e]

3. Intelligent AgriTrade to Abet Indian Farming KalpitaWagaskar,NilakshiJoshi,Amiya KumarTripathy,GauriDatar,SurajSinghvi,Rohan Paul

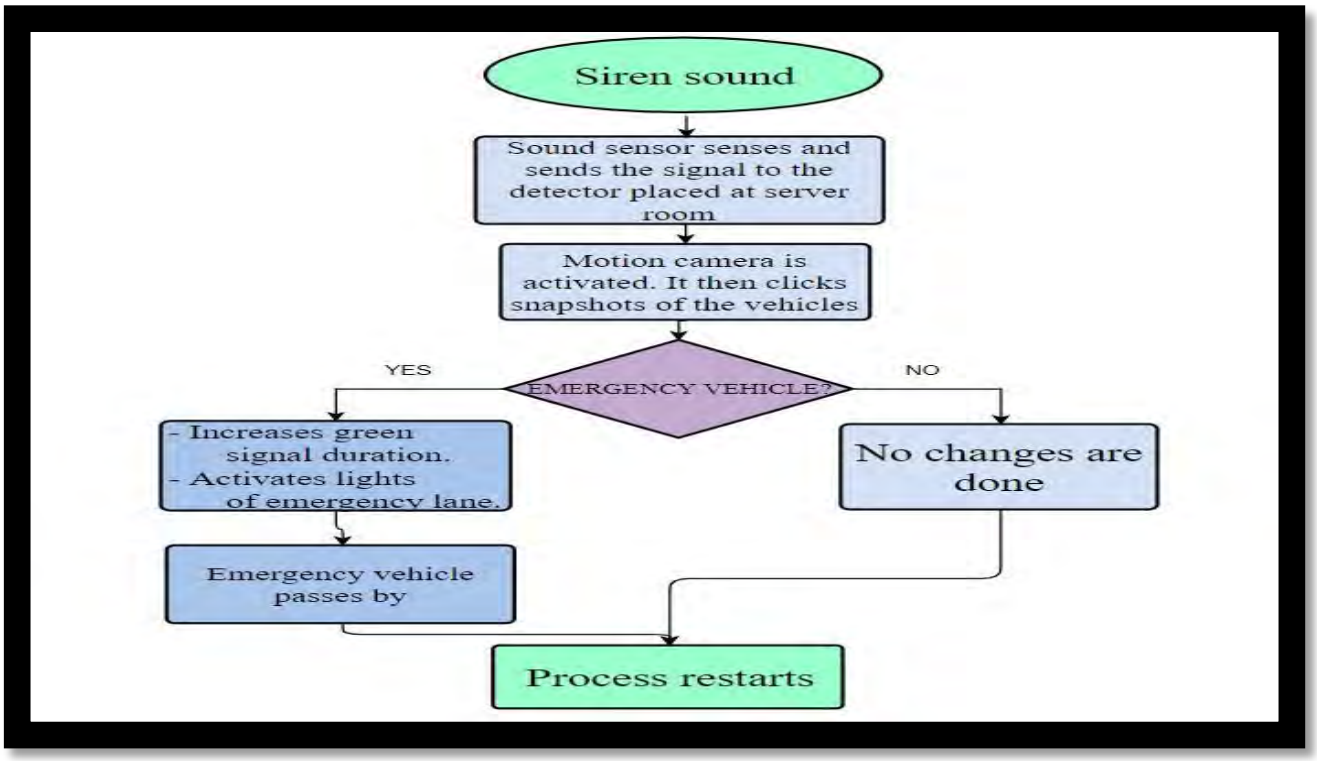
Traffic management for emergency vehicles

ABSTRACT: **- Technology can bring back anything but not life. Life is valuable. Everyday hundreds of people lose their lives due to delay in the reach of emergency services. The major contributor is traffic. Due to traffic the emergency vehicles like ambulance, fire engines, etc. are unable to reach in time at the destination, due to which the common people suffer. This project aims to solve the problem of traffic for the emergency services to some extent. The focus is to provide a way for the emergency vehicles on a busy road using artificial intelligence. A neural network is trained to identify emergency vehicles.**

KEYWORDS: -tesseract, neural network, artificial intelligence

WORKING: - All the emergency vehicles have a particular siren which is used when they have to reach a particular destination. The sound sensing sensors will be placed at about 2 km before the traffic signals and then at a distance of 1 km motion sensing cameras will be placed. When sound sensing sensor senses the sound of the siren it will be detected by the detector placed in the server room, and then the cameras will be activated. The camera will then start clicking pictures and send it to server room. Using these clicked pictures, neural network will identify whether the vehicle is an emergency vehicle or not. Once it is confirmed that the vehicle is an emergency vehicle, LED lights will start glowing in the emergency lane; a message will be displayed on the indicator as "Make way for the emergency vehicles" and the duration of green signal will increase. This will help the emergency vehicle to reach its destination in the shortest time it could.

FLOWCHART:-



Li-FI(DATA through ILLUMINATION)

Shashakt Jha , Shubham Joshi ,Utkarsh Kulkarni

Thakur College of Engineering and Technology

Mumbai INDIA

shashaktjha@gmail.com shubhamjoshi1303@gmail.com Kulutkarsh2000@gmail.com

*Abstract:-***This paper provides the basic idea, principle and applications of Li-Fi.**

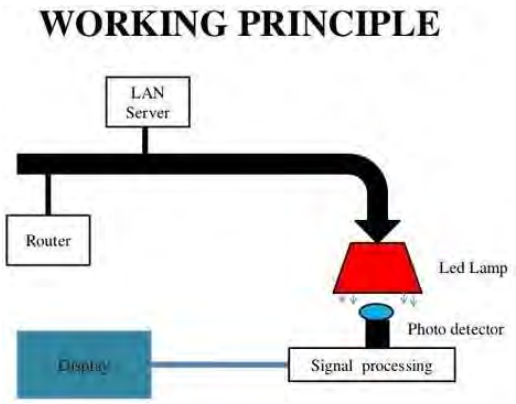
I. INTRODUCTION:-

From the beginning of mankind,
COMMUNICATION is one of the basic necessities for our survival as we share important data and Information with each other. But with ever increasing knowledge our current means (of data sharing) seems to be incompatible with the demand in terms of SPEED, RANGE and SECURITIES. In 2011, at TED Global Talk, Professor Harald Haas presented a new medium for sharing data via "VISIBLE LIGHT COMMUNICATION". This Technology is termed "LIGHT FIDELITY" or "Li-Fi". Li-Fi is a technology for wireless communicationBetween devices using light to transmit data andPosition. When an electrical current is applied to anLED light bulb, a constant stream of light is emitted from the bulb which we see as illumination. As LED bulbs are electronic devices the input current can be controlled and therefore the output light can be modulated at extremely high speed, this can be detected by a photo-detector and converted back to electrical current. Hence data can be transmitted at high speeds.Li-Fi is bi-directional full duplex and high speed i.e. data is transmitted and received at the same time resulting in a really fast and reliable network.

II. WORKING:-

Idea was very simple that if the LED is "on" then
The digital 1 can be transmitted and if the LED is "Off" then the digital 0 can be transmitted. Leds can be switched on and off very quick. For transmittingData this way all that we require is LED's andController that code data into LED's.

Parallel data transmission can be done by usingArray of LED's or by using red, green, blue LED'sTo alter light frequency with the frequency ofDifferent data channel. The data transmission is doneThrough binary codes which involve switching onLED can be done by logic 1 and switch off usingLogic 0. The encoding of information in light canTherefore be identified by varying the rate at whichThe LED's flicker on and off to give strings of 0'sAnd 1's. Visible light communication is this methodOf using rapid pulses of light to transmit information wirelessly



Visible Light Communication (VLC) is aData communication Medium, which uses visibleLight between 400 THz (780 nm) and 800 THz(375 Nm) as optical carrier for data transmission and Illumination. Fast pulses are used for wirelessTransmission. Communication system componentsAre:
1. A high brightness white LED which acts as aCommunication source
2. Silicon photo diode which shows good responseTo visible wavelength region.

LED illumination can be used as a communicationSource by modulating the LED light with the dataSignal. The LED light appears constant to the humanEye due to the fast flickering rate. The high data rateCan be achieved by using high speed LED's andAppropriate multiplexing techniques. Each LEDTransmits at a different data rate which can beIncreased by parallel data transmission using

LEDArrays. Many different reasons exist for the usage ofLED light in spite of fluorescent lamp, incandescentBulb etc. which are available.

Advantages over current systems:-

- Easy to install.
- Data Transfer rate for internet application is higher.
- Bandwidth of Li-Fi is 10000 times widerThan conventional radio waves system.
- It provides high amount of security as dataCommunication path is the Line of sightItself. Hence, this will avoid unwanted accessof Li-Fi signal by unauthorized users sittingIn the neighboring rooms.
- The Li-Fi devices consume low power forOperation and hence can be used in Iota(Internet of things) applications.
- Uses Optical spectrum and hence avoidsAlready crowded RF (Radio Frequency)Spectrum.
- There are no health concerns in Li-Fi basedSystems as it uses clean, pure visible band ofThe electromagnetic spectrum.
- There is a drastic energy reduction in Lighting industries that use Li-Fi based Devices.

III. DISADVANTAGES OR LIMITATIONS:-

- Internet can be used only where light sourceIs available.
- as light cannot penetrate optically opaqueBarriers, its range is limited.
- Though it draws low power, in order to availLi-Fi Internet services, light needs to be keptON throughout the day and also at night asInternet is the need of the hour.

IV. CONCLUSIONS:-

Li-Fi is a new emerging technology that mayReplace Wi-Fi in coming 10 years and opens newWays to explore an learn about data sharing with aVery high rate of transfer.

References:-

1) TED talk by Professor Harald Haas
2) IJSER International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013 1041.

Basics of Networking using TCP/IP model for Web based Applications

Mr. Anurag Singh , Mr. Harsh Singh , Ms. Pradnya Saval
Department of Computer Engineering
Thakur college of Engineering and Technology Mumbai,India
anurag.r.singh1234@gmail.com , harshsatishsingh1174@gmail.com , pradnyasaval@gmail.com

Abstract—**The TCP/IP model has been used forvarious data communication applications for several years based on a set of Protocols that govern the working of the Computer Networking System. This protocol suite is named TCP/IP due to the TCP (Transmission Control Protocol) and IP (Internet Protocol) protocols which are acknowledged as the most prominent protocols for digital communication. TCP/IP model is also referred to as the Internet Reference model or simply Internet model. TCP/IP Model is a four-layered data communication model that was developed by DARPA, an agency of the United States Department of Defence. The applications are based on the client-server model which contains a requesting node and a servicing node like sending an email, sharing files via the internet, visiting a website, etc. It consists of the Application layer at the top, followed by the Transport layer, Internet layer and Link layer at the bottom. The transmission of data is done by breaking the data into packets, frames, bits and then reassembled at the receiver's end where each layer has a particular task in the handling of data and to perform faithful communication between the nodes. In this article, an explanatory overview of the structure, working of the TCP/IP model, the importance of the essential protocols and its association of the model with the web world are explained.**

Keywords— Computer Networks, TCP/IP reference model, Web Based Applications.

I. INTRODUCTION

The TCP/IP model is a well-defined combination of protocols that define the process of data transmission from one node to another. This was developed earlier than the OSI model.[5] TCP/IP model provides complete assistance in assembling and presentation of data received by the receiver. This model is regarded as stateless because it considers every request as a new one and has got no relation with the previously made requests.[3] The model with the help of its 4 abstraction layers manages the assembling of data received, and provides the receiver with the desired message.

It is classified into four layers that help in the process of packetization, addressing, transmission, routing and receival at the other end. This has been the backbone of Computer Networking since its induction, it is found among applications that were produced a few years back such as Voice Over the Internet.[5] This feature is possible in this model with the help of Stream Control Transmission Protocol.

II. LITERATURE SURVEY

This was a result of the combined study of Defense Advanced Research Projects Agency in 1960's.The initial model of TCP consisted of datagram transmission as well as routing.. With the modifications being brought into this, it is considered as the most comprehensive suite for the general networking system. This technology was introduced to the market by IBM and AT&T during 1984.

A. Architectural structure

The working of this model is related to the end to end principle which states that the nodes in a network must be associated with Simplicity of communication. The modern day changes have caused this principle to be altered several times.

B. Comparison Between TCP/IP And OSI Model :

The difference between these two models begins with their classification of layers. The OSI model has a separate Application Layer that handles the interaction with user, the Presentation Layer that takes care about the encryption or Decryption of data and a Session Layer that manages the allocation of ports and sessions establishments. The TCP/IP model has classified all this under one head known as the Application Layer. The transport layer and Internet Layer remain the same for both the models. OSI models have a Data Link Layer and a Physical Layer towards the end which is converted to Link Layer for TCP model. Moreover, OSI model can be considered as a theoretical model whereas TCP/IP model is actually used for connecting to Computers (nodes) and Communication over a network. OSI model is utilised to Understand the working of the network. TCP provides with the feature of ease of connection whereas, OSI model helps the standardization of Routers or other Hardware Devices. TCP model actually helps to form a connection between two or more nodes.

III. FLOW OF DATA TRANSMISSION

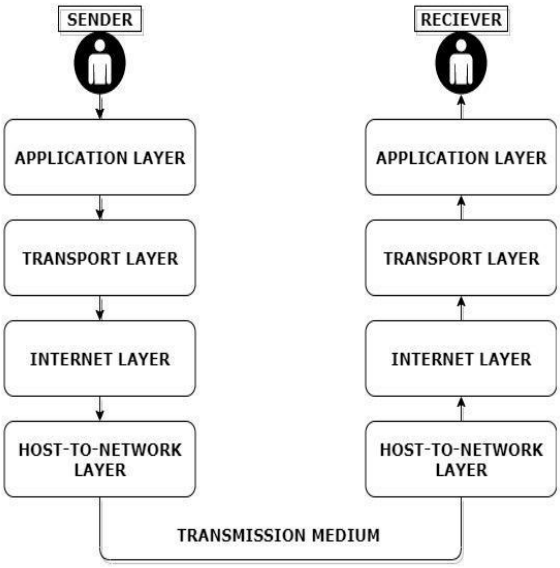


Fig 1.Flow of data Transmission

When a user sends a service request then it passes through four layers and finally through the transmission medium. Based on the Transmission medium the data can be passed in the form electrical signal (Analog or Digital) using cables, as a light signal by using fiber optic mode of transmission, etc. [5].The transmission of signals through cables is further divided into 2 types of viz. Baseband and Broadband. In the Baseband form of transmission, the data is directly transferred in the form of digital
pulses while the broadband uses the Modulator/Demodulator (Modem) circuit to convert the digital signal into analog and superimpose the data signal with a carrier wave and then transmit it through the channel [5].

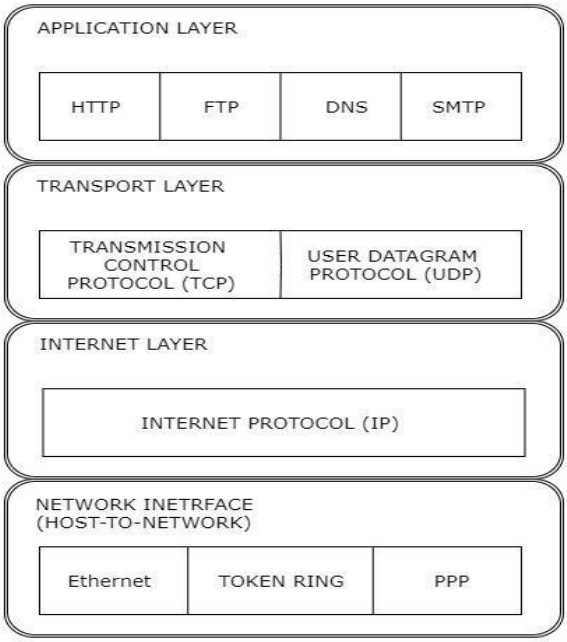


Fig 2. TCP/IP Protocol Suite

The brief Explanation of the 4 layers of the TCP/IP model :

A. Application Layer

The application layer is the interface between the user and the following layers. This layer enables the users to share data through the network and deals with the data generation, translations, formatting,
data compression and data encryption. This Application layer provides the following services:

Protocol	Used for
HTTP (Hyper Text Transfer Protocol)	Transmitting hypermedia documents e.g. HTML files, generally used between web browsers and web servers.
FTP (File Transfer Protocol)	Transmitting files over network amongst two or more hosts.
Telnet (Teletype Network)	For remote login and to execute functions remotely through the network which is bi-directional and provides a text based communication through terminal
DNS (Domain Naming System)	Naming the connected nodes in the network to make it easy to remember, also it translates the name to the IP address using a DNS server.
SMTP (Simple Mail Transfer Protocol)	Sending and receiving electronic mails (e-mails) on the network.

B. Transport Layer

The transport layer is considered to be one of the most important layers as this creates host-to-host communication. Even if the data reaches the receiver it is useless if it does not provide the detail of the process which requires that data. This issue is fixed by assigning port numbers to the processes running on the host machine. When the server responds to the request it encapsulates the port number at which the data needs to be delivered.

The most widely used protocols are shown in the below diagram:

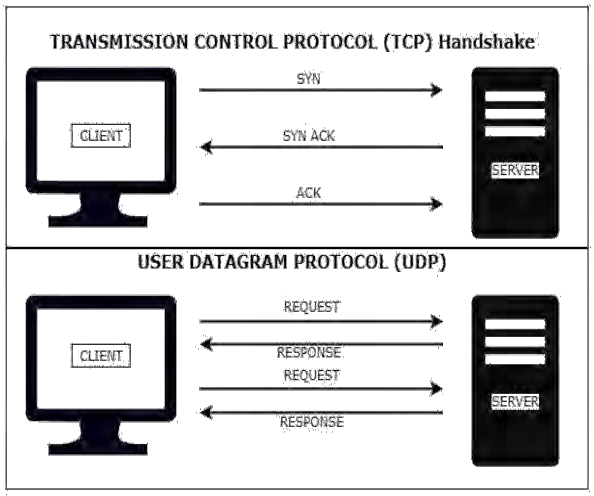


Fig 3. TCP and UDP protocols.

a) UDP (USER DATAGRAM PROTOCOL)

- It is a connectionless communication model.
- It doesn't keep track of the data transmitted
- Fails in identifying the retransmission condition.
- It is faster as compared to TCP protocol as it does not require a handshake before data transmission.
- It just verifies the checksum and accepts the data but no acknowledgement is sent back to the sender.
- The data block sent by the UDP protocol is called as Datagram.

b) TCP (TRANSMISSION CONTROL PROTOCOL)

- It provides a communication channel between the applications on client and server machines.
- TCP is a connection-oriented based protocol to enable host-to-host communication.
- In TCP a handshake is done between the two communicating nodes.
- TCP provides reliability, as it checks for errors in the received message and sends an acknowledgement if everything is alright within a specified time or else the sender resends the data.
- This enables retransmission of data as sender can make decisions based on the acknowledgement status [4].
- The transmitted block of data by TCP is called as Segment.

C. Internet Layer Or Network Layer

Network layer is completely responsible for inter-network communication. It doesn't rely on any physical topology of the network. The Datagrams/Segments are converted into Packet by adding sender and receiver's IP (Internet Protocol). This Packets are routed via different routers and reach the destination. The most commonly used hardware device working at this layer are routers.

D. Network Interface Layer

At this layer the Physical Address/MAC (Media Access Controller) Address of the source and destination are added to the packet. Also, the CRC (Cyclic Redundancy Check) bits are added as trailer also known as FCS (Frame Check Sequence) which help in error detection. The ready block to transfer is called Frame and now converted into bits of information and transferred through the medium. The devices working at this layer are Switches, Bridges , Hubs, Cables, Connectors, etc. These include the physical topology of the network and also these devices combine to form a single network.

IV. TYPES OF ADDRESSES

A. IP Address

IP Address is the unique id of a node in the network and beyond. Every device from the Internet layer has an unique IP address to give that device an identification. IP address is 32bit long i.e. 4 bytes. It has 4 8-bit numbers separated by '.' . These groups of binary numbers form decimal numbers in range 0-255. [1] For e.g. 192.168.10.5

B. Mac Address

MAC address is 6 bytes unique identification number assigned by the manufacturer to the NIC (Network Interface Card) .The 6 bytes are separated by-'. For eg. 00-1a-3f-f1-4c-c6

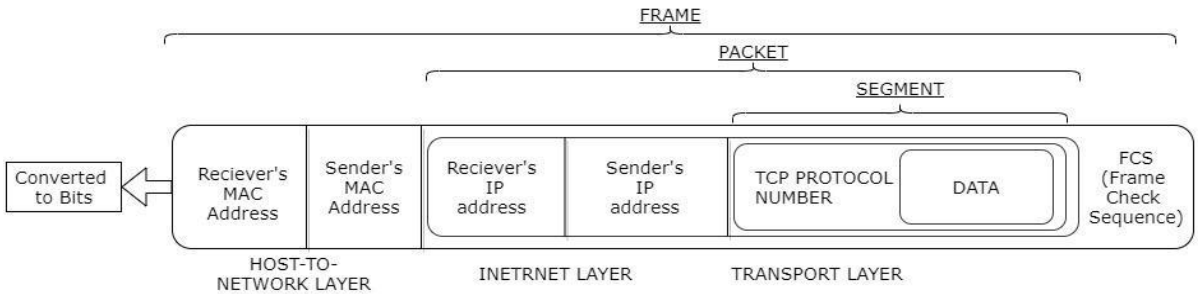


Fig. 4. Each layer Encapsulation

V. APPLICATION OF TCP/IP MODEL IN WEB

From loading a simple webpage onto the browser to uploading a picture on social media everything is dependent on the connection with the server and the network between you and server. Generally, the request sent by the browser is HTTP when requesting a web page. But before that, there's a middle-man called DNS server. We usually open a website by its URL (Uniform Resource Locator), this is a name given to a particular IP address that identifies the server. We cannot connect with the server without this.

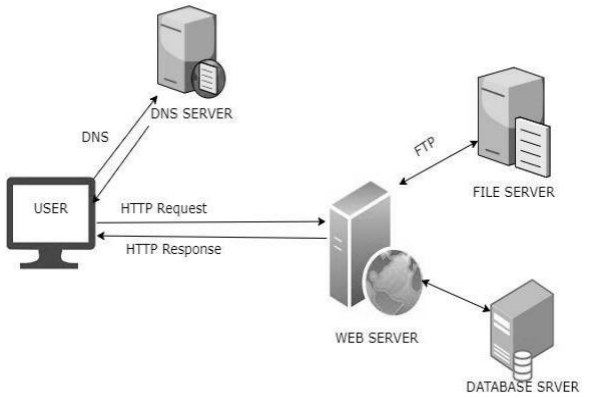


Fig. 5. WEB based Client- Server Communication model.

When a user sends a page request using URL, initially it acts as a DNS client and sends a DNS query to the DNS server requesting for the IP address allocated to the requested domain name. After receiving the IP address the connection is initiated. Then the host machine and Web server tries to do TLS (Transport Layer Security) handshake which basically runs on Transport Layer. This sets up the connection with the web server. Then all the data packets are sent to the server through routers using various routing algorithms and various other network devices. It is not necessary that all packets take the same route and reach the server at the same time. The reassembling of the segmented data is done at the receiver's end and then interpreted by the Web Server. If the user request is for data from a database or if the user is requesting images or some file so the web server connects to the database server using ssTCP/IP network or file server using FTP protocol respectively. The next step is to process the data received from database and file storage and send a HTTP response to the waiting user.

VI. CONCLUSION

The applications of TCP/IP protocols are vast, the model is used according to the convenience of the user group. The basic requirements of this modelscales down to a set of hardware devices and a particular software required for the sharing of data packets over a network[3]. Due to this factor TCP/IP model is implemented on all computer platforms. This model is responsible for data transmission, for all perspectives. The web browsers are being accessed every now and then, this shows that whenever we are trying to access some data from the internet or provide some data to the Internet, we are following the TCP/IP model.

While we request for some data from the internet, we have created a condition where two nodes exist, the requesting node and the server which responds and serves to our request becomes the servicing node. This technology model came into existence in 1960's

[3] but its applications are visible today as well. The model has gone through several changes according to the requirement of industry but the working still remains the well.

References

[1] Edmund Lam, TCP/IP Fundamentals [Online], University Of California ,(1999) ,available: http://www.sfisaca.org/download/lam.pdf

[2] Adomas Svirskas "The TCP/IP Reference Model" , [Online] St. Cloud State University (1999) available: http://www.mif.vu.lt/~adam/courses/npij/scsu-mcs426-fall-1999-3.pdf

[3] Pranab Bandhu Nath , Md.Mofiz Uddin," TCP-IP Model in Data Communication and Networking"[Online] ,Volume-4, Issue-10,(2015), available: http://ajer.org/papers/v4(10)/N04101020107.pdf

[4] Shreya Gangane, Prof. Vinit Kakade "Base of the Networking Protocol – TCP/IP Its Design and Security Aspects "[Online],Vol. 3, Issue 4, (2015) available: http://www.ijircce.com/upload/2015/april/144_Base.pdf

[5] Behrouz A. Forouzan "Data Communications and Networking" [Offline],Fourth Edition,McGraw Hill,(2007)

Effectiveness of 'S' in HTTPS

Saurabh Shetty , Shubham Pandey
Department of Computer Engineering
Thakur College of Engineering and Technology (TCET)
Kandivali East, Mumbai-400101
saurabh0503s@gmail.com, Shubham2sp2pandey@gmail.com

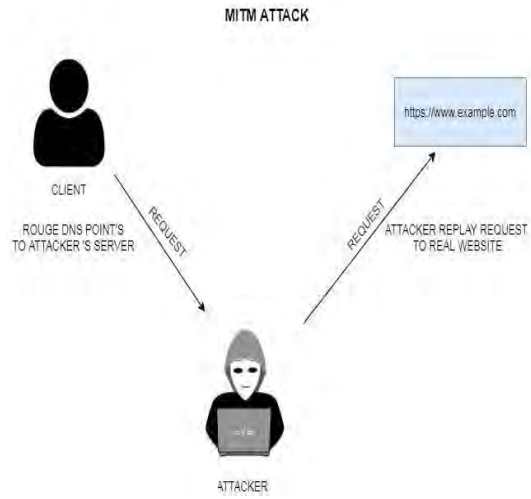
Abstract—**with the world becoming more connected than everwith the advent of internet, it has never become more important than ever to preserve the private and public data of fellow denizens.**
Many people are under the false influence that accessing a website which has HTTPS (Hypertext Transfer Protocol Secure) automatically makes them invulnerable from attacks and other third party data thieves, which isn't entirely true.
Most of the website uses SSL (Secure Sockets Layer) to ward off the Authorized detection and decoding of confidential data over the network. HTTPS protocol is used for communication between web browser and a websites server. While effective, HTTPS is not perfect and has its own flaws and loopholes, which a experienced hacker or any institute can manipulate the data, using SSL-MITM (Man in the middle) and other techniques. To be secure, while surfing the internet VPN(Virtual Private Network) ,SSL,HTTPS,ARP(Address Resolution Protocol) TABLE,ANTI-SNIFFING,ARP WATCH are advised to be used[1].

I .INTRODUCTION

The HTTPS is responsible for verifying/validating the communicating end points while providing confidentiality and integrity ensuing communication. However, as with most security solutions, it does not come with its own drawbacks. There are many research about protecting secure data, such as and focused on Wi-Fi Networks and human factors, secure the online retrieval of certificates by confirming certificate's fingerprint out-of-band. Identity-based cryptographic algorithms which do not require certificates i.e. a party's public key the party's identifier (e.g., fully qualified domain name or IP address).
This paper addresses the problem of the drawback of HTTPS and proposes 3
Different modus operandi to recover the loopholes of the protocol. [2]

II.SSL MITM Background

The following steps to decode the SSL using MITM.
1. Spoofing hacker-machine into victim-machine by Informing the gateway-router.
2. Spoofing hacker-machine into gateway-router by Informing the victim.
3. Enable packet routing feature on hacker-machine.
4. DNS Spoofing to enforce the victim to connects 443 portAt Attacker machine.
5. Forwarding fake certificate to victim.
6. Communication establishment with the victim using fakeCertificate.
7. Connecting with the HTTPS website by using genuine Certificate gathered from the HTTPS website. .
8. Capturing the packets transmission between devices.
9. Decrypt data. [3]

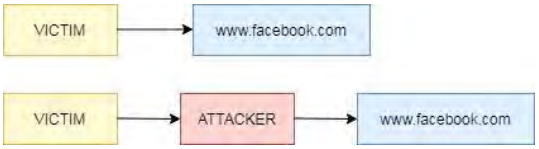


III. Offensive Way

Here we simulated a scenario in which we are trying to get Facebook password of a user. To do this, we need to understand 3 terminologies i.e. Victim IP Address(192.168.1.10), gateway-router IP address (192.168.1.1) and attacking machine IP address (192.168.1.20).Also, ensure that victim and attacker machines are in the same network. MAC Address of an attacker suppose is AA-BB-CC-00-11-22. The steps to perform the attack are:

1.Enable the packet routing in your Linux operating system. Command is echo 1 > /pros/sys/net/ipv4/ip_forward
2.Make sure that hacker spoof itself as a Gateway router. suds arpspoof –t 192.168.1.10 192.168.1.1
Thus, Victim believes that Gateway MAC Address is AA-BB-CC-00-11-22, which is originally attacker IP Address. As a result, all packets of data is been transmitted through attacker machine.
3.Attacker machine has to inform the Gateway router and spoof itself as a victim machine.

Suds arpspoof –t 192.168.1.1 192.168.1.10 Here hacker is performing Man in the middle attack.[4]



4. Whenever victim visits Facebook, the packets are been transmitted through attacker machine. Hacker can use any sniffing tools available such as Wire shark, Ethereal or website.
Thus, Attacker can capture username and password. And can store the credentials for later use.
5. Here we save a credentials in store data file and decode the data as follows:
Slump –r store data –kwebmitm.crt > out
Hacker can easily read data sent by victim in clear format. [6]

IV. DEFENSIVE WAY

1. Apply ARP WatchThis help to prevent attack as ARP Watch maintains the list of changes done in ARP value. This is applied at client side. This is useful as it can trigger and record the event. Disadvantages arise due to compatibility issue such as software cannot be installed on Router.
2. Apply Anti Sniff It can identify and indicates some IP Address that is, running a packet capture tool. The admin can block the IP Address.
3. Apply static ARP Advantage of using static ARP technique is admin is configuring switch rather than client. Since switches have "Port Secure" feature. Do note that some of the switches do not have "Port Secure". [5]

V.CONCLUSION

Scammers, phishers, hackers and other entities are always on the hunt to steal your public and private data .Our study proves that HTTPS, while a highly effective and secure way of protecting your data isn't entirely perfect. There a various ways which a user must follow to ensure his online privacy and safeguarding of personal data is being assured.

VI. FUTURE SCOPE

Confidentiality, Integrity and Authentication of the Website is ensured if the website is using HTTPS rather than HTTP. Most of the HTTPS servers are vulnerable to trivial MITM attack. Thus, to ensure the security which hasMonitoring AlertingAnti-SniffingTechnology must be integrated with HTTPS. Such as merging the technology contents of VPN, SSL, HTTPS, ARP TABLE, ANTI-SNIFFING, and ARP WATCH.

References

[1] Thawatchai Chomsiri, "HTTPS Hacking Protection", Proceeding of 21st International Conference on Advanced Information Networking and Applications Workshops, 2007, AINAW '07 , IEEE CS Press, Niagara Falls CANADA, May 2007, pp. 590 - 594.

[2]D. Naylor, A. Finamore, I. Leontiadis, Y. Grunenberger, M. Mellia, M. Munafò, K. Papagiannaki, and P. Steenkiste. The Cost of the "S" in HTTPS. In Proc. ACM CoNeXT, 2014.

[3] Keith W. Ross, James F. Kurrose., Addison Wesley Longman, "Computer Networking: a top-down approach featuring the Internet", Inc., USA, 2004.

[4]Angela D. Orebaugh, and Gilbert Ramirez, "Ethereal Packet Sniffing", Syngress Publishing, February 2004.

[5]Xia H., and Brustoloni J., "Detecting and Blocking Unauthorized Access in Wi-Fi Networks", In Proc. Networking'2004, IFIP, Lecture Notes in Computer Science, 3042:795-806, Springer-Verlag, May 2004. <http://www.cs.pitt.edu/~jcb/papers/net2004.pdf>

[6]Thawatchai Chomsiri, and Preecha Noiumkar, "Decoding HTTPS secure data on LAN", Draft Technical Report, Faculty of Informatics, Mahasarakham University, Mahasarakham Thailand, January 20

Ocean cleaning- Plastic

Ms. Vidhi Punjabi Ms. Janhavi Tawde,Ms. Munazza Gaoher Ms. Urvi Shukla
Department of Computer
Thakur college of Engineering and Technology
Mumbai, India
vidhpunjabi@gmail.comtawdejanhavi@yahoo.communazzagaoher@gmail.comurvishukla20@gmail.in

***Abstract*-The environment pays a heavy price for our fast fashion habits. Every year 300 million tons of plastic is produced out of which 8 million tons of plastic goes into the ocean. This is causing cross linked impact on the planet by not only affecting the sea life but also due to absorption PCBs and DDTs it poisons the food chain. Diverse solutions have been proposed ranging from source reduction to ocean based clean up. Our paper focuses on eliminating the plastic waste from the oceans by using the ocean currents. An installation of U-shaped screen channels moving with the help of waves similar to that of plastic which concentrates the floating plastic to a center point. This concentrated plastic can be extracted and recycled. Thus, making it profitable.**

Keywords—buoyancy, micro plastic, floater

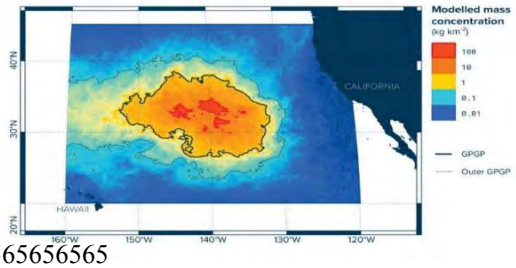
I. INTRODUCTION

Suppose a water pipe in your kitchen has broken up and the water is flowing continuously. What will you do? Close the main water supply while keeping the spilled water as it or cleaning the spilled water while the tap is still running? Sounds absurd to execute any one of the above ideas? It's the same with the oceans. We cannot focus only on the methods to reduce the plastic production while keeping our oceans unclean. Both the actions should go on simultaneously. Reduced plastic production and ocean cleaning should go hand in hand to improve the quality of sea water and enrich the marine biodiversity. There are various methods undertaken to reduce plastic usage. But the latter one is completely ignored. Our paper focuses on the need to clean the oceans while providing different, innovative ideas to do so.Plastic is cheap, durable and flexible. Hence, we use it in almost everything. But when the same plastic ends up in the ocean it has drastic impact on our life. Its imbalances the entire food chain by wiping out the marine species. The hazardous effects cannot be ignored. So, we thought of an idea to clean the oceans. The idea is to make the oceanic currents do the work. Because why move through the oceans when oceans can move through you? By installing our system in the ocean which consists of U-shaped channels with screens below. It is influenced by the wind, waves and the currents. It is a flexible, movable system which is self-supporting, receiving energy from sun, wind and waves. It can bring the plastic to the center of the system. This accumulated plastic can be collected by sending the ships to the area where the plastic is concentrated. The location of the concentrated plastic a be tracked down with the help of satellites by tracing the position of our system.

II.THE NEED FOR CLEANING

8 million tons of plastic founds its way to the oceans. Due to which 5 plastic gyres are found in the oceans. The great pacific garbage patch is one of the major plastic gyres situated in the north central Pacific Ocean. Its estimated area is 1.6 million square kilometers which is about thrice the area of France. 1.8 trillion tons of plastic is floating in this patch. The estimated weight of plastic is 80,000 tons in this gyre. The center of the patch has high density of plastic while the boundaries have low density of plastic. The density of plastic from center to the boundary ranges from 100skg/sq.km to 10skg/sq.km respectively. It would take thousands of years to clean such large amount of garbage using the conventional methods. But with our idea we can clean it by 50% in a span of five years [1].

Fig.1 (a): The great pacific garbage patch [2]



35656565656565

Fig.1 (b): mass concentration of plastic in the patch [1]

The high-grade density plastic used to make various materials is broken down by sun, waves and curious critters forming micro plastics. Micro plastics are plastics having size less than 5mm. they are categorized into two types a) primary micro plastics b) secondary micro plastics. The primary.

Micro plastics are intentionally produced like plastic pellets, industrial abrasive, exfoliates etc. The secondary micro plastic is formed by erosion of large sized plastic especially marine debris.

III. EFFECT ON MARINE LIFE

Every minute a truck full of plastic is unloaded in the oceans. The effects caused on our marine ecosystem cannot be ignored. The plastic started accumulating in the sea water from 1960s. Millions of marine animals are killed due to plastic debris every year. 700 species of marine life have gone extinct and the number will increase exponentially if we continue this trend [3].

Plastic is mistaken as food by the marine animals. This results into their death due to starvation because by eating plastic their stomach feels full and so they don't eat what they are supposed to. This leads to malnutrition. Sea turtles presumably are victims to this problem. Blockage in their digestive system is one of the major reasons for their percentage decrease in population. Marine mammals which swim closer to the surface of the water are found caught in marine litters like fishing gears, nets, plastic bags etc. In a study it was found that 388 sea lions entangled in plastic debris 3) Sea birds die each year due to consumption of micro plastics while catching fishes. 98% of albatross studied are found to have ingested some kind of plastics in different amounts [3]. The carcasses of whales and dolphins are frequently found on the shores with large amount of plastic ingested. These all incidents have led scientists and researchers to find ways to clean ocean.

Fig.2 (a): sea lion entangled in plastic. [3]

IV. PROPOSED SYSTEM

Instead of using conventional ideas to clean the oceans like vessels and nets we have proposed this system which deploys the ocean currents. It is heavily influenced by three factors-winds, waves and currents. The principle behind it is simple-to create coastline where there are none, concentrate the plastic and take it out.

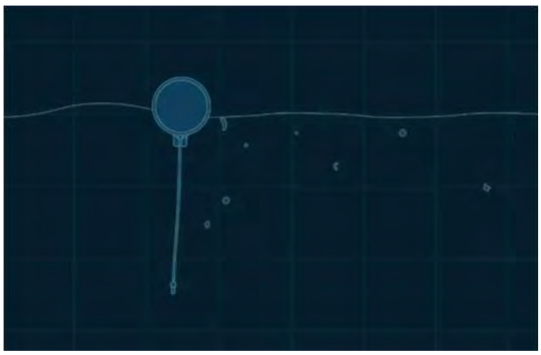


Fig.3 (a): floater above the surface with anchor and skirt below. [1]

The system consists of a 600m long floater and 3m deep skirt attached below. The floater provides buoyancy to plastic and prevents plastic from flowing over it. The skirt which is a screen prevents smaller particles from escaping underneath. The model has had undergone many modifications in the design. Instead of fixing the system to the sea bed which was proposed earlier, an anchor up to 600m deep is used. This provides stability whilst mobility. This allows to gather plastic in more efficient way. The system primarily depends on currents, winds and waves. The currents move the plastic and the system both. But the winds propel the floater only as it

lies above the surface of the water and as most of the plastic lies beneath the surface of the water. This pushes the system faster than the plastic thereby collecting garbage more efficiently. [1]

The skirt of the system is deeper in the central p art rather than the edges. As impenetrable skirt provides downward flow which helps the marine life to escape from below it. The current forces it to adopt U-shaped structure which concentrates the plastic like a funnel. The drag force generated by the skirt acts like a stabilizing force allowing the system to re-orient itself when wind changes direction. The system has solar power lights, anti-collision system, cameras, sensors and satellite antennas. The system communicates its position at all time and continuously gather performance data. Then a support vessel is sends to the live position where it extracts the plastic with the help of nets. [1]

The fleet of system consists of 60 such systems. If executed it will manage to clean the Great Pacific Garbage Patch by 50% in 5 years and by 90% until 2040. [1]

The biggest challenge is to survive storms. The key to the survival is flexibility. While designing the structure various cases were considered that can occur in many years. The model has undergone many tests to withstand the forces of the ocean. The system has been built in such a way that it follows waves because it's free floating.

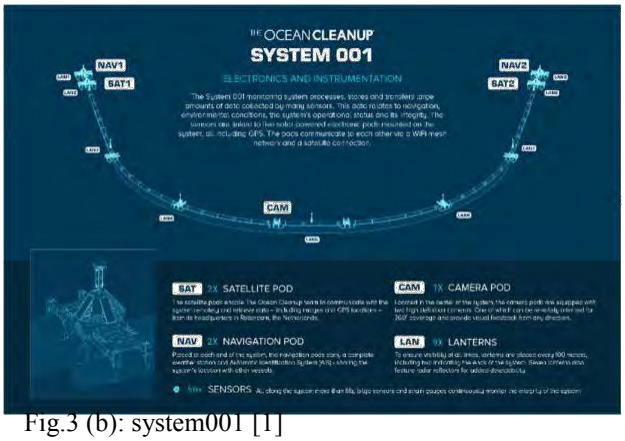


Fig.3 (b): system001 [1]

The plastic on the surface of the ocean or up to a depth of few meters deep can be cleaned using this system. But the plastic on the sea bed needs a different approach altogether for undertaking the cleaning process. Vast amounts of funds, manpower and emissions can be saved by executing thedf

V. FUTURE SCOPE

Waste production will not stop but will only minimize if we care enough to take preventive measures. Likewise dumping plastic in the oceans won't reduce to zero. We need to think many different and diverse solutions to clean the oceans. It is a really challenging task to clean the ocean. Conventional methods would take many years as well as large capital. By executing it and learning from mistakes various modification can be done. The system should be ocean friendly and feasible. The system is a pioneer in the advanced cleaning operation of the oceans. We need to invent and upgrade such models for dirt-free ocean.

VII. CONCLUSION

we, humans are solely responsible for creating this havoc which is causing serious threat to our marine life and ecosystem which indirectly affects us. To redeem the devastation caused by plastic to our oceans we proposed this
Idea. It will not eliminate plastic completely from the oceans as majority of plastic lies on the seabeds. But this measure will mitigate the amount of plastic present in our oceans. Thus,
Reducing the chances of damage to the marine life and ecosystem.
It is a breakthrough in the way of cleaning the oceans. The proposed approach is a basic effort to clean the plastic garbage that floats across our beautiful oceans thereby giving our

VI. REFERENCES

[1] www.theoceancleanup.com

[2] www.timesofindia.com

[3] www.onegreenplanet.org

[4]www.marinedebris.noaa.gov/sites/default/files/publications-files/TM_NOS_ORR_30.pdf

Advance Surveillance Security Using AI and Face Detection

Mr. P.U.Tembhare ,Dhanashri Adgurwar, Sharayu Bagmare, Anup Katre, Sanket Harinkhede ,Vaibhav Rahangdale

Department of Computer Technology, Priyadarshini College of Engineering

Nagpur, India

Abstract- **Face detection and recognition from an image or a video is one of the most interesting, and well known topic in biometric research as it deals with security and privacy purpose. The human face plays a very crucial role in the social association and offers each and every individual a unique identity face recognition technology has broadly attracted attention due to its massive application value, and market potential, such as real time video surveillance and fraud detection. It is widely recognized that the face recognition plays a prominent in surveillance system as it doesn't require the participation of an object. We used image set algorithm with the aid of OpenCV and Python programming development and having a pre-existing database of faces. The module distributed in three sub modules: Detection module, Training module and Recognition module.**

Keywords: Face detection, Face Recognition, OpenCV.

I. INTRODUCTION-

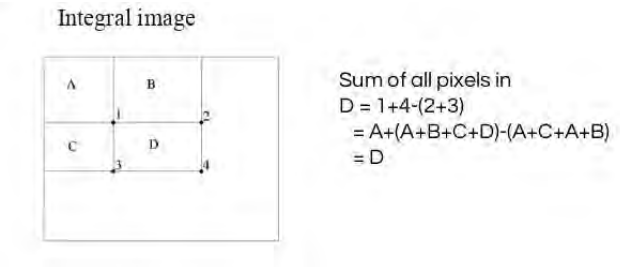
Face recognition technology was commonly observed as a something straight out of science fiction. But over the past decade, this groundbreaking technology has not just become viable, but also it has become widespread. Face detection and recognition is technology which is used to recognize an individual from a video or photo source. The pioneers of facial recognition were Woodrow Wilson Bledsoe, Helen Chan Wolf and Charls Bisson. In the 1960s the concept of face recognition was introduced by Woodrow Wilson Bledsoe. Bledsoe developed a system called as RAND tablet that could classify photos of faces by hand. RAND tablet is a device that people could use to input horizontal and vertical coordinates on a grid using a stylus that emitted electromagnetic pulses. The system could be used to record manually the coordinate locations of various facial features including the eyes, nose, hairline and mouth. These set of data could then be inserted in a database. Then, when the system was given a new image of an individual, it was able to retrieve the image from the database that most closely resembled to that individual. Ever since then recognition system is being improved and optimized constantly, the technology becomes gradually mature and is more and more widely used in human daily life. There are several industries benefiting from this technology. Law enforcement agencies are using face recognition technology to keep communities safer from impostors. Retailers are preventing crime and violence. Airports are improving traveler's convenience and security. Nowadays mobile phone companies are using face recognition technology to provide consumers with new layers of biometric security. In this paper, we propose a face detection and recognition system with the aid of python along with OpenCV. This system contains three modules which are detection module, training module and recognition module. The detection module recognizes the face which gets into the field of vision of the camera and saves the face in the form of an image in JPG format. Then the training modules trains the system with the aid of Haar cascade algorithm which was proposed by Paul Viola and Michael Jones. This method consists of four steps-

1. HAAR FEATURE SELECTION-

First step is to collect the Haar Features. [8]A Haar feature considers adjacent rectangular areas at a specific position in a detection window, which adds the pixel intensities in each area and calculates the difference between these additions.

II. CREATING INTEGRAL IMAGES-

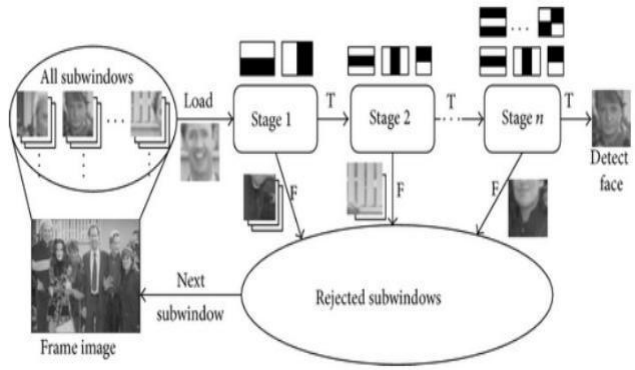
Integral Images are used to make this procedure swift. Most of the calculated features are irrelevant.



III. ADABOOST TRAINING-

A concept called Adaboost which selects the best features and then trains the classifiers which is used. This algorithm creates a sturdy classifier using a linear combination of weighted simple weak classifiers.

IV. CASCADING CLASSIFIERS-



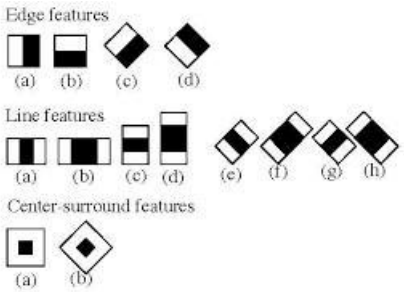
The cascade classifier comprise of a number of stages, where each stage is a group of weak learners. These weak learners are simple classifiers which are known as decision stumps. Each stage is trained by using a Technique called boosting. Boosting provides the potential to train a highly precise classifier by taking the weighted average of decisions which are made by the weak learners.

Finally, in the recognition module the fundamental components of the face from the new video are extracted. Then those features get compared with the list of elements stored during training and the ones with the best match is found and name of the person recognized is displayed. This monitoring system fulfills the basic needs of face detection and recognition system, also takes the cost into consideration to ensure the pervasive mode as economical as possible. Furthermore, it can also be combined with real-time analysis algorithms.

V. PROBLEM DEFINITION- FACE DETECTION

The main function of face detection is to determine [1] Whether human faces appear in a given image or not, and [2] where these faces are located at given image. The expected outputs of this step are patches containing each face in the input image. Hence to make face recognition more robust and efficient to design, face alignment are performed to justify the scales and orientations of these patches. [3]Once a face is detected, we have to extract a face signature. The finest facial recognition systems are aligning the faces before extracting the face signature as it makes the face signature improved. Aligning is the process of rotating and scaling the face in order to bring all of the key features into the same places (eyes, nose, mouth etc.).This is important because it allows the function generating the face signature to work on much cleaner input, and makes recognition perform much better.

VI. FEATURE EXTRACTION-



After the face detection step, human-face patches are removed from images via feature extraction phase. If we use these patches for face recognition directly, it may have some limitations; first, each patch usually holds over 1000 pixels, which are too enormous to build a robust recognition. Second, face patches may be taken under different illumination, with different face expressions, and with different camera alignment, and may suffer from occlusion and cluttering background. To overcome these drawbacks, we perform feature extractions to do dimensionality reduction, information packing, salience extraction, and noise cleaning. After this step, we transformed face patch into a vector with fixed dimension or a set of fiducially points and their corresponding locations. We can include feature extraction either in face detection or face recognition as per survey form some literatures.

VII. MODEL TRAINING

Simply put, model training is the process of taking several face representation belonging to the same person, and creating a classifier that can recognize more instances of face portrayal from the same person. The most favored approach is to use a linear SVM.

VIII. FACE RECOGNITION

After formulating each face patch in the above steps, the last step is to recognize the identities of these faces. For face recognition, a face database is required to build. [4]For each individual, some images are taken and their features are removed and stored in the database. Whenever an input image came, we perform face detection and face extraction and after this we compare features of each face with stored database.

We differentiate two applications of face recognition: identification and verification. Face identification means given a face image, we want the system to tell who he / she is and identify the individual ; while in face verification, given a face image and on the basis of a assumption from the identification, we want the system to tell true or false about the guess.

IX. DESCRIPTION OF TOOLS

In this section, the tools and approach to implement and calculate face detection and tracking using OpenCV are detailed.

X. OPENCV

[8]Open Source Computer Vision Library is widely known as OpenCV. It is a library of programming methods and mainly focused at the real time computer vision which is developed by Intel. The library is cross-platform and it mainly aims at real-time image processing. Originally the library is written in C and this C interface makes OpenCV portable to some distinct platforms such as digital signal processors. Wrappers for languages such as C#, Python, Ruby and Java (using JavaCV) have been developed to uplift adoption by a wider user. However, since version 2.0, OpenCV includes both its traditional C interface and a new advance C++ interface. This new interface seeks to eliminate the number of lines of code essential to code up vision functionality as well as eliminate common programming errors such as memory leaks (through automatic data allocation and deallocation) that can occurs when using OpenCV in C. The majority of developments and algorithms in OpenCV are now developed in the C++ interface. [9]It is much tougher to provide wrappers in other languages to C++ code as opposed to C code; since the other language wrappers are generally lacking some latest OpenCV 2.0 features.

VII. PROPOSED SOLUTION-

When image quality is taken into consideration, there is a plethora of aspects that influence the system's correctness. It is significantly important to apply distinct image pre-processing techniques to standardize the images that you supply to a face recognition system. [2]Most face recognition algorithms are very responsive to lighting conditions, so that if it was trained to detect an individual when they are in a dark room, it probably won't recognize them in a bright room, etc. This problem is mentioned to as "lamination dependent", and there are also many other issues, such as the face should be in a very steady position within the images (such as the pixel coordinates of eyes have to be same), compatible size, rotation angle, hair and makeup, emotion (smiling, angry, sad, etc.), position of lights (to the left or right, above, etc.). This is why it is foremost

to use a good image pre-processing filters before applying face recognition. You should also do things like removing the pixels around the face which are not in used, such as with an elliptical mask that shows the inner face region, not the image background and hairs, since they modify more than the face does. For clarity, the face recognition system presented in this paper is [1]Eigen faces with the aid of grayscale images. The paper assert how easily is to convert color images to grayscale, and then to apply Histogram Equalization as a very easy method of automatically standardizing the brightness and contrast of the facial images. For better output, you could utilize color face recognition, or employ more processing stages such as edge enhancement, contour detection, motion detection, etc. Also, this code rescales images to a standard size, but this may change the aspect ratio of the face. OpenCV utilize a type of face detector called a Haar Cascade classifier. An image, which can come from a real time video or from file, the face detector examines each image position and classifies it as "Face" or "Not Face." Classification assumes a stiff scale for the face, say 50x50 pixels. Since faces in an image might be larger or smaller than this, the classifier runs over the image multiple times, to search for faces across a range of particular scales.

VII. CONCLUSION-

In this paper we have developed a system whose main objective is to detect and recognize the faces for providing high security. The dataset is created for storing the images. The images of the person are then trained and defined before recognizing. The Haar cascade algorithm is used for identifying the faces. OpenCV offers pre-trained Haar cascade algorithms. By using tremendous amount of training data (in the form of images) the classifier begins by extracting Haar features from each image. Not all features are useful for identifying face. Adaboost a machine learning algorithm is used for selecting best features. In the coming future, as technology advances, more advance features will be added to the system.

Reference-

[1] M.A.Turk and A.P. Pentaland, "Face Recognition Using Eigenfaces", IEEE conf. on Computer Vision and Pattern Recognition, pp. 586-591, 1991

[2] Ahonen, Timo, et al. "Recognition of blurred faces using local phase quantization." Pattern Recognition, 2008. ICPR 2008.19th International Conference on. IEEE 2008.

[3] CH. Y. Lu, CH.SH. Zhang & F. Wen (1999), "Regional Feature based Fast Human Face Detection", J. Tsinghua Univ. (Sci. and Tech.), China, Vol. 39, Pp. 101–105.

[4] H. J. Jiang (2007), "Research on Household Anti-Theft System based on Face Recognition Technology", Nanjing University of Aeronautics and Astronautics, China.

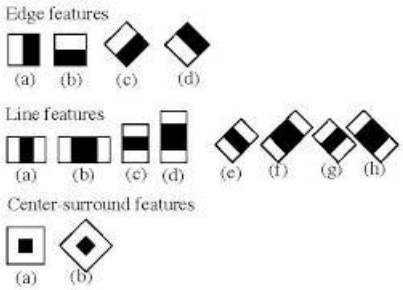
[5]https://www.youtube.com/watch?list=PLQVvvaa0QuDfKTOs3Keq_kaG2P55YRn5v&v=OGxgnH8y2NM

[6]https://www.youtube.com/watch?list=PLQVvvaa0QuDfKTOs3Keq_kaG2P55YRn5v&v=OGxgnH8y2NM

[7]https://docs.opencv.org/3.4/d7/d8b/tutorial_py_face_detection.html

[8]https://pythonprogramming.net/haar-cascade-object-detection-python-opencv-tutorial/

[9].Open Source Computer Vision Library Reference Manual-Intel [Media]



Intelligent Essay Scrutinizing System

NupurDeshmukh[1],RohitKahar[2],DiptiPangam[3],VivianBrianLobo[4]
[1,2,3]U.G. Student, [4]Assistant Professor
Department ofComputer Engineering
St.JohnCollegeofEngineeringandManagement, Palghar
[1]nprdshmkh@gmail.com,[2]rohitkahar33@gmail.com,[3]pangam.dipti.99@gmail.com,[4]lobo.vivian27@gmail.com

Abstract—**withincreasing numberof students optingfor varioustypes ofEnglishproficiency tests,thereisaneedto expandthe horizonsofmethodoftakingexaminations—from traditional pen–papermethodtogivingthem online.Various examinations likeTest ofEnglish asaForeign Language (TOEFL),GeneralRecords Examinations(GRE),International English LanguageTesting Service(IELTS),etc.haveseenan increaseinthenumberofpeopleattemptingforthem. Allof thesecompetitiveexaminations haveawritingskillssection, i.e., essaywriting.Onecan'tjustgrade or assesslarge number of essays quickly. In addition, insufficienttopicknowledge canaffectthe accuracyin grading orassessingessays.This iswhere aninterfacethatis unbiasedandatthe same time saves timeof grading orevaluatingessaysisneeded.Thisstudyemphasizeson solvingthe abovementioned problem by developingawell-built interfacethatcanassist peoplefor gradingessays. Weprepareto dothis byconsideringfeatureslike quality ofcontent,count of sentences,words,etc. presentinanessay anditsrelevance of topic.**

Keywords—text analysis,grading,naturallanguage toolkit, datasets

I. INTRODUCTION

Nowadays, manyinstitutesandschoolswhoconduct languagetestsand/orproficiencytesthaveessaysinthem andassessingtheseessaysmanuallyiswearisome forany personasittakes alargeamountoftime.Inordertohandle eachessayappropriately,weneedanintelligentsystemthat isadvanced intermsofproviding aquicksolution. We requirefast, effective,andaffordablesolutionsforautomated gradingofstudent-written essays.Analyzingahuman languageandinterpretingitcouldbedifficultforamachine tounderstand. Thisisbecauseasentencecanhavemany differentmeaningsorinterpretations,andasystemmightn ot interpretitright.Togetthis right,wefeedthesystemwitha largenumber ofdatasetsandclassify thislargesetoftextual entitiesintosmalldiscretecategoriessothatitbecomes easy togradethem.Thisstudyfocuses ontextanalysis andtext miningapproachandgradesessaysaccordingly. Anattempt ismadefornaturallanguage processing usingnatural languagetoolkit(NLTK).

The remainderof thispaper is organized as follows. SectionIIfocuses onthereviewofliterature thatisrelevant tointelligentessayscrutinizing system.SectionIIIdescribed theproposedsystem.SectionIV concludesthestudy.

II. REVIEW OF LITERATURE

Inaresearchconducted byZestandWojatzki[1],the systembuiltbythemconsidered onlytask-independent features,andthesefeatureswereforeseeable toexecutewell irrespective ofwhattaskitwastrainedon.Theexperiments wereperformedwithacontemporaryessaygradingsystem onGermanandEnglishdatasets. Therewascategorization betweenfeaturesintotask-dependentand-independentones

Andevaluationofdifferenceingradingperfectionbetwee n analogousmodels.Theydeterminedthattask-independent modelsexecuted weresuperior forboththelanguages tested, butthederived losseswerecomparatively highonthewhole.Thesystem tookasetofgradedessaysandpre-processed themusing tokenization,part-of-speech (POS)tagging,stemming,and syntacticparsing.Featureextrication hadalistoffeatures (eithercomplete orreducedsetoffeatures), andthey extractedcorresponding featurevaluesfromoccurrences.A machinelearning(ML)algorithm thenlearned a representation modelofessayqualityfromextricated features.Inthenextindependentstep,thelearnedmodelw as putinordertoevaluateessays.Theevaluation metric produced avalueforagreement between humangoldstandardandmachine grading.Thedrawback foundintheir workwasthatthemachine couldnotprovetobemuch efficient. Arafat and Raihanuzzaman[2] intheirresearch usedlinear regressionand randomforest algorithm.The mainobjective of theirresearchwastodevelopanintelligent cloud-basedsystemthatcouldautomaticallygradeanessay aswellasoffervaluablesuggestions. Theyusedpython's NLTKfornaturallanguageprocessingtoextractfeatures likewordcount,sentence count,andsentence towordratio, andattemptedtoobservedifferent patternspresentinessays toscorethem. Theyextractednoticeable featuresfromessayscreated fromdataset withnecessaryformationandagainused supervisedMLmodelsto buildanartificialsystemthatcould score further usergivenessays andalsomakesuggestions. They implementeda MLagent that is trainedbylinearregression algorithmonextractedfeatures topredictscores andcalculatedcosine distancetodetermine similarhelpful essaysandrecommended thoseessaystousers.Moreover, theydevelopedtheirsystem tosuggestawriter necessary correction oftheirmistakesandwritingpatterns.Afterthe prediction ofscoresweremadethesystemsuggested few improvementstothetestessaysfromcollection ofessaysby whichausercouldgothrough thosesuggested essaysjustto makethatperson's nextessayimprovedthantheprevious one.Thesystem maybeefficient,butasperourobservations, thesystem couldbebiasedtocontext.Ifthetestessay didn't match thetrainingessays, thenthesystemmightscore erroneously. Also,theremaybesomekeytextspresentin essaysthatareirrelevanttoessaysthatmightchangethe scoreofessays.

Harika *etal.*[3]intheir researchmade use ofsequential forward featureselection algorithm. The goal of their research wasto make the systemgrasp onto one of the human languages and problems present in them.

This enabledthemtoknowaboutautomatedsystems andtestit withMLalgorithmstocreateastableinterfacetoserveour motive.Theprocess ofalgorithmwassimilartoitsname itself.Itsequentiallycommencedwithnullset andproceeded furtherbyadding onefeatureaftertheotherintheset.The inputdatasetwastakenfromKaggle. Everyessaywas repeatedly processedwithstepslikestopwordremoval, tokenization,and stemming.Their researchalso included 'casefolding' method.Itisamethodtoconvertnon-uppercaselettersto uppercaseletters.Afterthedatawas gathered and pre-processed, various forms of techniques wereusedtoextricatevarious noticeable featuresor indicatorsfromessayssothatarepresentationmodelcould betrainedforanticipating thescore.Thenotableattributes werenumberofparagraphs, sentences,words.Inaddition, attributeslikenumberofstopwords,stemwords,andunique wordsexistedinanessay.

Theywereoftheopinionthatautomatically evaluating essays will helpwriters toknow theirownlevelof proficiency. However, therewasavastscope of improvement,especiallywith thedifficultyfor languages aroundtheworld. Eveninthissystem,someofthecritical grammatical errorswerequiteimpossibletonotice,andthey mighthavegoneunnoticed. However,amajordrawback of their research was the algorithm, i.e., sequentialforward featureselection, asitdidn'tassure higher performance. Furthermore, ifallcandidate featureswereequally relevant forthetaskathand, removing these features wouldbe harmful.

Ramalingam*etal.*[4]intheirresearchmadeanattempt to create a system that could be precise in providing feedbackaswellasassessing performance.Themainaimof theirresearchwastocreate atime-efficientsystemthatsaves time of students as well as teachers. To build a robust system,theyconsideredattributessuchaslanguageelegancy, grammatical andsyntacticprecision, vocabularyand classification ofwordsused,essaylength,anddomain information. Fortrivial featureslikecountofwords, sentences,andparagraphs,theyusedtextmining.Toidentify POSpresentinanessay,theymade useofpython library NLTK.

Moreover, forcheckingthecorrectness ofspelling, another pythonlibraryenchantwasused.Toidentify the relevancy oftopic,theauthorsselectedabestessayfora particulartopicandextractednounsfrom them,whichserved askeywords forthattopic.Thedatasetthatwasusedwas Kaggle. After grabbing the dataset and performing operationson them,thescorewascomputed.

Also,humanandmachine resultswerecompared inthe formofgraphs. Thisresearchlackedsemanticandsyntactic features.Everysooften,therearemanyotherfactors that could playaroleindeterminingthescore ofanessayfor increasingitsaccuracy.

ShankarandRavibabu [5]intheirresearchused sequentialforwardselectionalgorithm.Thechiefaimwasto resolvemanualcheckingcarriedoutbyhumansthat required lotoftimeandenergy.

Theyusedfeatureextraction techniquewithmost common features like word repetition, sentence authorization, andspellchecker.Thus,basedonautomated essayevaluation andbytheuseoffeatureextrication methods,ageneralessaywasevaluatedfirstandthenthe finalresultwasproduced. Theybegantheirstudywith "model generation" usinginputdatasetwithalreadygraded essays inthem.Valuesforextractedfeatureswerecomputed, andrubricswereplanned. Everyessaywasrepetitively processedwithstepslikestopwordremoval,tokenization andstemming.Later,inputtest data wasexposedtothe same stepsasthetrainingdata.Thesecomputations were performedforallfeatures,andresultsweregeneratedwithin arangeof1–5.They hypothesizedthatagoodevaluationfor anessayscorewouldinvolvearangeofattributes suchas language fluencyandagility, phrasingandvocabulary, structureandorganization, andcontent.Moresystematic modelwasextracted usingtheabovefeatures. Theproposed algorithmwassuitableforidentifyingthecorrectness and relevanceofanessayforagiventopic.

They collecteddatafrom fifteen students,which was consideredasatestdataset.Asperresults,itswasinterpreted thattherewashardlyanydifference betweenmanually generatedscoreandsystemgenerated score.Hence,the proposedsystemwasvalid.Theirresearchshowed that software-based essaygradingwillhelpwriterstoknowhow welltheycanwriteanessay.

However, therearemanysections ofadvancements, especiallywith thedifficultyforlanguagesaroundtheworld. Eveninthissystem,someofthegrammatical errorswere impossibleto record, and they might have gone undiscovered. Thedrawbackoftheirsystemwasthatthey contemplatedaccuracyastheonlyestimation.

III.PROPOSED SYSTEM

Figure1showstheproposedsystemarchitecture of intelligent essayscrutinizing systemTheproposedsystemis ananticipated modelthatistrainedontwoinputs(i.e., examiner's modelanswerandstudent'sanswer).The proposedsystemconsists offourmodules,namely preprocessing module,mappingmodule,feedbackmodule, andvalidationmodule.

Examiner's modelanswerandstudent'sanswer,which aretheinputs,arepreprocessedinapreprocessingmodule to mapwordsindatasets.Thepreprocessing moduleinvolves componentslikeparagraphandsentencealignmentfollowed bytokenization (i.e.,dividesentencesinsmallparts),which inturnincludeswordalignmentandstopwordremovalsuch as "the","a","an","in".

Next,welookforexactword matchingwhichcan bein exactlythesameorderorinascattered order.Afterthis, stemming isperformed,i.e.,reducingwordstoitsrootform. Notethattheretwoprocessesareperformedin mapping module.

Thefeedback module checks forsimilarcomputations betweenexaminers' modelanswerandstudent's answerand assignsascoreaccordingly.

Moreover, thestudentanswerischeckedmanuallyby humanandgradedscorebasedonhumanjudgment.

Next, thesystem computed score iscompared withhumanratedscoreandscoredaccordingly.

ThesystemutilizesNLTKlibraryinpythonandHe wlett FoundationdatasetofCagletoacquirebestpossibleresults.

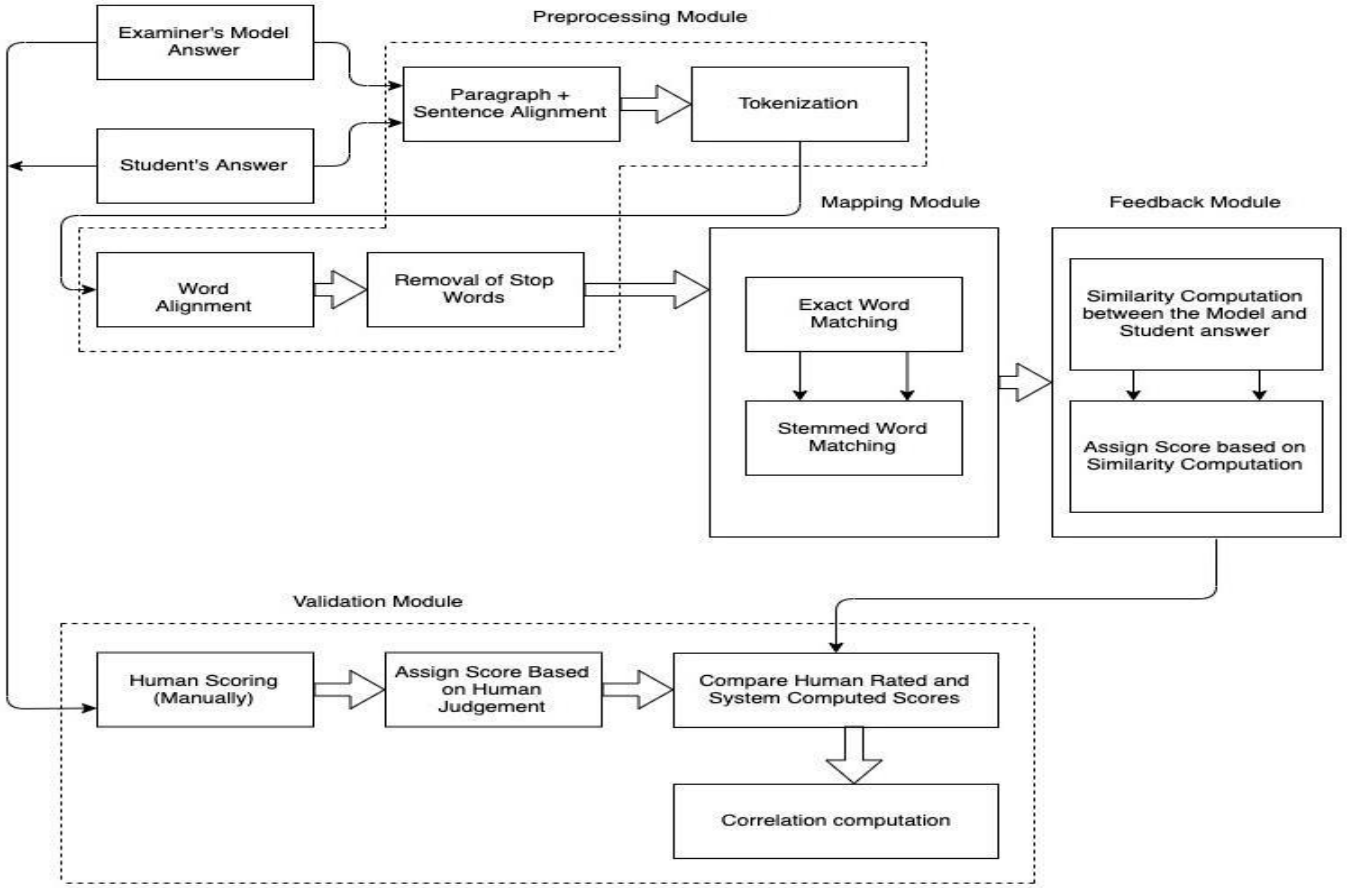


Fig. 1. Proposedsystemarchitectureofintelligentessayscrutinizing System

IV.CONCLUSION

The proposed system ofautomaticallygrading essays looks positive andcanbeimproved withbetterfeaturesets thatcapturedetailsaboutsentence structure andarrangement ofideas ratherthansimplyfocusing onsurface featuresand keywordsinessays.Goingforward,therewillbesystems thatwilleffectively gradeessaysinthemosthuman-like manneraspossible.

V. ACKNOWLEDGMENTS

Theauthors wouldliketothankDr.G.V.Mulgund, Principal,St.JohnCollegeofEngineering andManagement (SJCEM),Palghar,Dr.RahulKhokale, Professor andHead, Department ofComputerEngineering,St.JohnCollegeof Engineering andManagement (SJCEM),Palghar,andthe entirestaffofDepartment ofComputerEngineeringof SJCEMformotivating ustowritethisproposedsystem paper.

References

[1] Zesch,T.,Wojatzki,M.,&Scholten-Akoun,D.(2015,June).Task- independentfeatures forautomatedessaygrading.*InProceedingsof theTenthWorkshoponInnovativeUse of NLPforBuilding EducationalApplications* (pp. 224-232).

[2] Arafat,A.,&Raihanuzzaman,M.(2016).Automatedessaygrading with recommendation*(Doctoral dissertation,BRAC University)*.

[3] Harika,Y.,Latha,I.S.,Sai,V.L.,Krishna,P.S.,&Suneetha,M. (2017).Automated EssayGradingUsingFeaturesSelection. *International Research Journal of Engineering and Technology*, (2017)March.

[4] Ramalingam,V.V.,Pandian,A.,Chetry,P.,&Nigam,H.(2018, April).Automatedessaygradingusingmachinelearningalgorithm.*In JournalofPhysics: Conference Series*(Vol.1000,No.1,p. 012030). IOP Publishing.

[5] Shankar,R.S.,&Ravibabu,D.(2019).DigitalReportGradingUsing NLPFeatureSelection.*InSoftComputinginDataAnalytics*(pp.615-623).Springer, Singapore.

Intelligent Chabot For E-Learning: A Study

Megharani Patil , Anand Khandare

Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
megharani.patil@thakureducation.org , anand.khandare@thakureducation.org

Abstract— **Nowadays India is moving towards digitalization in this context to provide any kind of help the use of Chatbots is very popular in many applications where systems provide an intelligence support to the user for solving their queries. To speed up the assistance many web based and mobile based systems are equipped with Chatbots that can interpret the user questions and provide the answers immediately. This work is all about prototype of a Chatbot in e-learning. This can provide support to college students to learn various courses. The design of the chatbot architecture should modelled in such a way that it can manage communication and provide the right answers to the student. Chatbot system should detect the questions and using the use of natural language processing techniques and available knowledge base provide the answers to student.**

***Keywords*—Chatbot, E-learning, NLP**

I. INTRODUCTION

Chatbot is a typical computer program designed to simulate an intelligent conversation with one or more users (human) through audio or text. Chatbots are set up as virtual assistants, represented visually in the form of a cartoon character, animation, or an avatar, and in some cases, a voice or text stream. Chatbots are designed to express human-like conversations through artificial intelligence. They can be programmed to operate as per a set of predefined instructions. The more a chatbot interacts and learns new things, the smarter it becomes. They can be accessed through various chat tools such as Facebook Messenger and other messaging applications[1]. They basically interact with the user just like a live person. Chatbots use the text-to-speech technology which helps them deliver responses in a simulated human voice, or even have text boxes that display the words being said. In eLearning, you can use chatbots as a means of communication. E.g you can enter text into a form field to which the chatbot will respond. If a system is set up with a speech recognition app, then you can vocally communicate with the chatbot. Chatbots are pre-programmed with a database containing a set of words, phrases, and questions. When a query is asked, it responds to the query by choosing a suitable word/phrase from the set of instructions in the database[2]. A Chatbot is a software that talks with an user. It is a virtual assistant able to answer a number of user questions, providing the required responses. Application areas of Chatbots are Health Care, Marketing, Educational, Supporting Systems, Cultural Heritage, Entertainment and many others. Chatbots can be developed both for industrial solutions and for research. Some example of chatbots are Apple Siri, Microsoft Cortana, Facebook M and IBM Watson. One of the most challenging tasks is the development of effective Chatbots: the emulation of human dialogues, which involves problems related to the NLP (Natural Language Processing) [3].

Chatbot has some problems. It is not possible to map all user requests, and the current Chatbots do not show remarkable performances because of the unpredictability of user thought during a conversation. The correct design of conversational flow plays an important role in the development of a Chatbot. For successful conversation, it is important to handle with all user requests and provide the required answers. This work is based on the knowledge base that can be used to interpret the intentions of the user and solve the problem of interpretation of sentences written by the user [4]. The Learning Connection at Chabot provides students and instructors with the support they need to achieve their learning and teaching goals as well as co-curricular activity. Chatbot technology can be considered an important innovation for e-learning. There are many innovative solution in filling the gap between technology and education. Chatbots creates an interactive learning experience for the students, like the one-to-one interaction with the teacher. Chatbot can also be used as a source of social learning, students from different backgrounds can share their views and perspectives on a specific matter while the bot can still adapt to each one of them individually. This technology can improve engagement among students and encourage interaction with the rest of the class by assigning group works and projects like teachers usually do. From testing the student's behavior and in order to keep track of their improvements, bots play an essential role in enhancing the skills of an individual student. They encourage a student to work by sending regular reminders and notifications. They provide a system for a personalized learning experience. Each student earns and absorbs things at a different pace. Using Chatbots is possible to adapt the speed at which a student can learn. They can also work as a support for teachers by identifying spelling and grammatical mistakes, checking homework, assigning projects and especially keeping track of progress and achievements of each student. . Chatbot

aims to be an e-Tutor for students. Some of the chatbot features in e-Learning are for

1. **Simulations:** Here you can use a robust version of a simulation bot where the bot plays a role of a character or guide. The bot can interact with the learner and guide him throughout the simulation.

2. **Chatbot as a source of reinforcement:** You can use chatbots for the reinforcement of learning. You can do this by pushing out information at certain intervals, in response to certain triggers.

3. **Social learning:** Social learning is another area where you can use chatbots. One of the benefits of social learning an environment is that the chatbot's interaction doesn't need to be limited to an individual. It can also communicate with a larger group.

Design considerations while designing chatbots for e-commerece.

1. If you are planning to use a chatbot as part of your eLearning, then you need to consider the nature and demographics of your intended audience. This will influence the syntax and tone of response of the chatbot.

2. You can use the existing course resourse books, handouts, and vidio clips of related media to boost the chatbot's database knowledge and vocabulary.

3. You can also use a list of Frequently Asked Questions to generate a chatbot's list of preprogrammed queries and responses. If you are planning to use an avatar or an animated character, ensure the chatbot's appearance is in sync with the audience it addresses. Pick a gender for your avatar and finally, don't forget to name your avatar.

II. RELATED WORK

Many Chatbot applications based on AIML are analyzed in particular an integrated platform which consists of a basic AIML knowledge is presented. In this project, Chatbot is called Tutorbot because it is functionality backing of didactics done in elearning environments. It contains some features as natural language management, presentation of contents, and interaction with search engine. Besides, e-learning platforms work is linked to indispensable services to web service. A continuous monitoring service has been created on e-learning platform servers which is another controlling machine[5]. Game based e-Learning tool used to teach digital forensic investigation. A Chatbot inside the learning platform helps students. A multiple-choice question based quiz is implemented for kinesthetic learners, and there is a pedagogical Chatbot agent that assists users. It provides easy navigation and interaction within the content. The Chatbot is implemented to be a pedagogical agent for the users, which is meant for discussions and help with the topics. It also acts as a navigation tool and can play video or use the advanced wiki if there are somethings to ask[6]. Investigated about the realization of intelligent agents on platform with IBM Watson technology. These agents in the form of Chatbots have to automate the interaction between the student and the teacher within the frames of Moodle learning management system. Watson is a cognitive system that combines capabilities in Natural Language Processing, analytics, and machine learning techniques. In this case, a Chatbot through the Facebook Messenger is realized to simplify communication between a teacher and a student: it could be arranged by acquiring Moodle test basis by Facebook Messenger Bot GUI Builder. A motivating example will be illustrated in the next session[7]. Using Bayesian theory to match the request of student and furnish the right response. In particular, Chatbot agent accepts to student's answers and extracts the keywords from the question using a lexical parser, then the keywords are compared with the category list database. The Bayesian probabilities are obtained for all categories in the list. Once the category is selected keywords are compared with the questions under the category using Bayesian probability theory. The answer to the question, which has the highest posterior probability, is then fed into the text to speech conversion module and thus the student receives the answer to his question as a voice response [8].

In literature, there are many approaches related to Chatbots, in particular on e-learning systems. From the beginning of the last decade the use of artificial intelligence as e-learning support has captured the interest of many researchers for its many applications. The web bot in an e-learning platform, to address the lack of real-time responses for the students. In fact, when a student asks a question on e-learning platform the teacher could answer at a later stage. If there are more students and more questions, this delay increases. Web bot is a web-based Chatbot that predicts future events based on keywords entered on the Internet. In this bot stores the questions and answers it on XML style language. This bot is trained with a series of questions and answers: when it cannot provide a response to a question, a human user is responsible for responding. In the last recent years some interesting research works can be found [9].

III. MOTIVATION

Teaching is a relational act based on communication and interaction, and chatbots have significant educational potential precisely due to their communicative ability through natural language. One of the reasons for the commitment to use chatbots in different economic activities such as customer care is increased process efficiency, for example, 24-hour care or specific information. This argument is also used in the case of education, as a chatbot can operate as a 24/7 support service, which can allow,

for example, teachers and service staff to avoid having to answer repetitive questions that can be easily resolved. The incorporation of chatbots in education has to be preceded by prior thought, whether its aim is educational or not. An institutional and organizational debate is needed to ensure functionality, feasibility and scalability within the institution. It is important to stress that the inclusion of chatbots will not replace teaching staff or administration and services staff but it can take over some of their tasks to complement and help them out. It appears to be feasible to think of a future where there is close collaboration between humans and machines, and, in the case of teaching, teacher roles may be distributed between both agents. The teacher of the future could become a sum of the human teacher and the AI teacher, with a complementary division of tasks. For example, the human teacher could take care of the creation of excerpts of learning materials, which could then be complemented or broadened by the AI teacher. The human teacher could take on personalized tutoring tasks, while the AI teacher provides uninterrupted support. The former could act in cases of dispute or in solving assessment appeals, while the AI teacher could answer FAQs, act as a virtual tutor to guide periodical tasks or redirect complaints and demands. This symbiosis would allow the human teacher to stop taking on more mechanical or repetitive tasks – which would be undertaken by the chatbot or AI teacher – and would consequently have greater availability to devote their time to more creative and high cognitive level tasks. We need to know, then, what the affordances of the different types of chatbot are to see how they operate as cognitive colleagues and not as potential threats. In the field of education, chatbots are used experimentally, making the most of the user interfaces and the boom in their popularity in different economic sectors for the promise of their benefits. However, it remains to be seen how they adapt to every context and how they are understood and rated by students, teachers, and administration and services staff. An example of an app that enables the creation of chatbots for education is SnatchBot. It is a free tool and has an interface that does not require programming knowledge. Its function will be determined by its creator's intention: either to answer informative queries (FAQ-type) or to act as a tutor to teach concepts and procedures. This presence of chatbots will also depend on the number and type of interactions that they can have with other chatbots or tools, either incorporating functions that are currently carried out by different apps or converging data so as to enable integrated actions. We are referring, for example, to the chatbot's ability to gather information from an individual's personal email, combine it with the calendar and the information available on a university's website, put the data together and be able to confirm a tutorial on the university's virtual campus. As has happened with other tools, chatbots may in the future undertake some assistant functions that have been done until now by specific apps. These affordances allow users to worry less about tasks closely related to memory (calendars, reminders, submission deadlines, instructions, etc). New advances in voice and emotional state recognition will end up smoothing these interactions.

IV. CLASSIFICATION OF CHATBOTS

Depending on the functions carried out by chatbots in education, we can classify them on the basis of the following tasks:

Administrative and management tasks to foster personal productivity: they provide personal assistance to students, aiding onboarding and personal productivity. Tasks include schedule or email management and task, submission deadline or assessment reminders. This uninterrupted personalization involves giving each student a rapid and personalized service, which takes pressure off academic services administration.

Taking care of FAQs: they provide a response to student FAQs regarding administration or learning concepts and contents. Unlike the first, they do not include personalization elements but student services in the form of FAQs. Tasks include information about admissions and enrolment, financial services, technical problems.

Student mentoring: they allow student mentoring during the learning process. They are able to respond emotionally, they monitor the student's understanding (cognitive control) and they can provide support and make suggestions to the student when needed. One of the main tasks is the provision and adaptation of contents. In this case, they are chatbots that enable educational programme contents to be generated and adapted, which are then sent straight to the user, taking their preferences into account.

Motivation: they contribute to exercising behavioural control by providing positive motivational reinforcement. This process means that students' retention is increased, which is especially relevant in online learning environments.

Practice of specific skills and abilities: they enable dialogues to be practiced in language learning, simulating conversations in contexts organized by level and with different roles and discourses.

Simulations: they simulate specific professional situations and can provide support for reflection or therapy. For example, in the field of healthcare, they can simulate patient treatment; in psychological care, they can simulate patient care; and in formal or social education, they can simulate the understanding of different school learning situations.

Reflection and metacognitive strategies: they help the students regulate their own metacognitive processes, they act as an expert classmate and they can provide support to aid learning. There are currently no examples of chatbots that offer this reflective function.

Student learning assessment: they can act as exercise assessors quickly and automatically. One example is automatic essay scoring, which gives feedback on mass courses based on automatic learning capable of analysing thousands of essays and giving an automated score. They also include tasks related to feedback to the student, who receives support thanks to learning feedback and the adaptation of the learning process at the student's pace and according to their needs.

V.ARCHITECTURE OF E-LEARNING CHATBOT

The architecture of composed of Front-End, Back-Office, Knowledge Base and E-learning BOT as shown in Figure 1. **Front-End**: it consists of different kinds of device like tablets, smartphones, PCs and so on.

Back-Office: It is used to manage operations that are not seen by end user. This module works in the background to better satisfy user demand. It handles business logic and data storage, working in collaboration with the knowledge base.

The Knowledge Base: It is a special type of database, where data is processed by a server, for the management of knowledge and information. Users representing all users of the application . Learning object is a collection of content items, practice items, and assessment items that are combined based on a single learning objective. **The E-learning BOT**: It is the main engine of the system. It is composed of Interaction Quality Tracker, Human- Computer Interaction Supervisor, Context-Aware Information Manager and Inference Engine.

- **Interaction Quality Tracker**: this module monitors interactions between users and Chatbot, evaluating conversation logs based on quality indicators and highlighting critical aspects of Human-machine interactions.

- **Human-Computer Interaction Supervisor**: it supervises dialogue, tracks interaction times, identifies ambiguous questions, recognizes nonconvergent interaction sessions, and indicates the need for community support if it is not possible to give a correct answer.

- **Context-Aware Information Manager**: this module allows to drive the dialogue based on contextual parameters. The goal is to provide a mechanism of dynamic and automatic invocation of information considering the context through the Context Dimension Tree [10].

- **Inference Engine:** It is designed to provide right answer to the user through Latent Dirichlet Allocation Algorithm and Workflow Manager. In first phase the definition of an ontology for the description of E-learning domain and the adoption of pre-existing ontologies is studied. The second phase is related to a workflow navigation module which can surf the ontology and select the more appropriate sentences. The description of the workflow can be achieved using Petri Net [11].

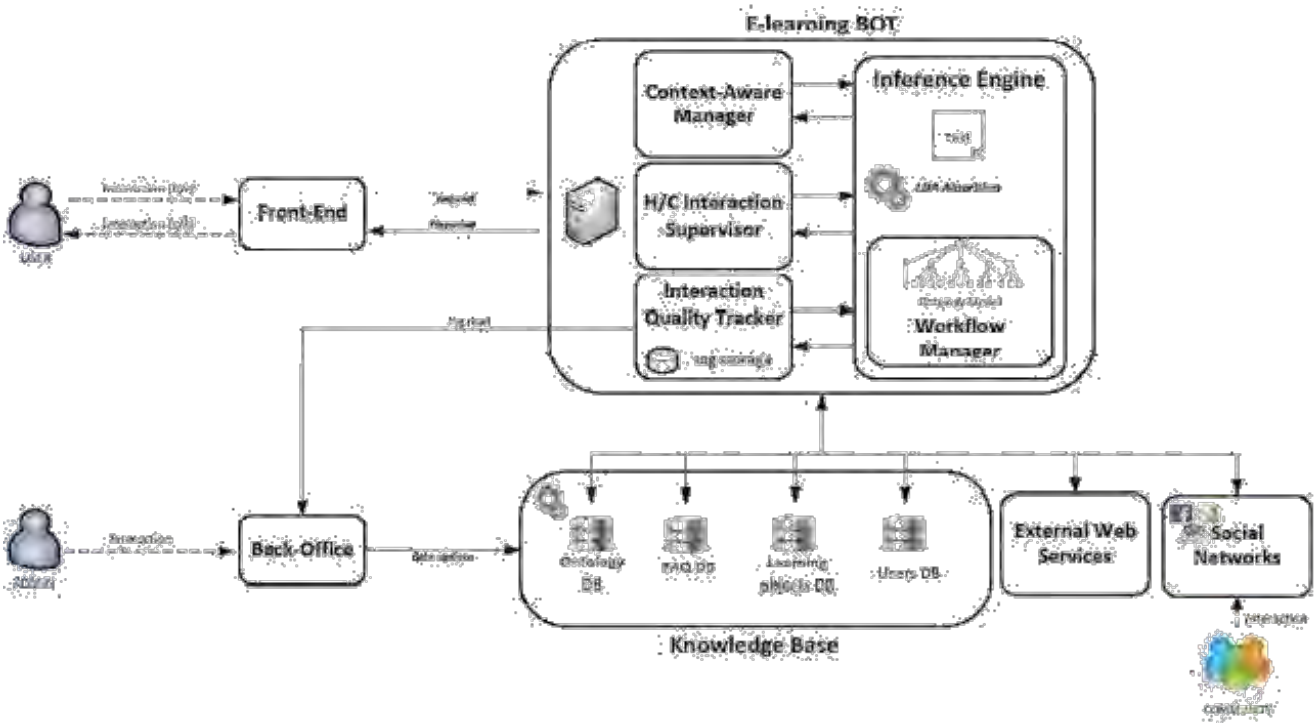


Fig 1: Architecture of E-learning chatbot

VI. CONCLUSION

This paper started with usefulness of chatbot in e-learning with features of chatbot and design consideration required to build a chatbot for e-commerce. There is summarization of related work and motivations towards the study of intelligent e-learning chatbots. It also included classification of chatbots followed by general architecture of intelligent e-learning chatbot.

VII. ACKNOWLEDGEMENT

My special thanks to our principal Dr. B.K. Mishra, Dean academic Dr.R. R. Sedamkar and HOD Dr. Sheetal Rathi for encouraging us to write this paper.

References

[1] A.D.Kane, "The Role of Chatbots in Teaching and Learning", In Rice, Scott; Gregor, M.N. E-Learning and the Academic Library: Essays on Innovative Initiatives. McFarland & Co. pp. 131–147, 2016

[2]Available at https://chatbotsmagazine.com/the-complete-beginner-s-guide-to-chatbots-8280b7b90/

[3] M. Yan, P. Castro, P. Cheng, V. Ishakian, "Building a chatbot with serverless Computing*," in Proc. the 1st International Workshop on Mashups of Things and APIs* Article No. 5, 2016.

[4] L. Lin, L. F. D'Haro, R. Banchs "A Web-based Platform for Collection of Human-Chatbot Interactions," *in Proc. the Fourth International Conference on Human Agent Interaction*, pp. 363-366, 2016.

[5] S. Satu, H. Parvez, and Shamim-AI-Mamun "Review of integrated applications with AIML based chatbot," *First International Conference on Computer and Information Engineering* (ICCIE), 2015.

[6] O. Nordhaug, A. S. Imran, Al. Alawawdeh and S. J. Kowalski "The forensic challenger," *International Conference on Web and Open Access to Learning* (ICWOAL), 2015.

[7] M. Niranjan, M. S. Saipreethy and G. T. Kumar "An intelligent question answering conversational agent using naïve Bayesian classifier," *International Conference on Technology Enhanced Education* (ICTEE), 2012.

[8] N. Nenkov, G. Dimitrov, Y. Dyachenko and K. Koeva "Artificial intelligence technologies for personnel learning management systems," *Eighth International Conference on Intelligent Systems*, 2015.

[9] M. Farhan, I. M. Munwar, M. Aslam, A. M. Martinez Enriquez, A. Farooq, S. Tanveer, and P. A. Mejia
"Automated reply to students' queries in e-learning environment using Web-BOT," *Eleventh Mexican*

International Conference on Artificial Intelligence:Advances in Artifical Intelligence and Applications, 2012.

[10] M. Casillo, L. Cerullo, F. Colace, S. Lemma, M. Lombardi and A. Pietrosanto, "An adaptive context aware app for the tourism," *The 3rd Multidisciplinary International Social Networks Conference (MISNC),* 2016.

[11] F. Colace, M. De Santo, S. Lemma, M. Lombardi, F. Pascale "BotWheels: a Petri Net based Chatbot for Recommending Tires," in *Proc. the 3rd edition of Special Session on Knowledge Discovery meets Information Systems: Applications of Big Data Analytics and BI - methodologies, techniques and tools (KomIS), within the 6th International Conference on Data Science, Technology and Applications (DATA),* 2017.

Implementing System for Predicting Heart Disease Risks Using Convolution Neural Network

Mr. Vikas Singh, Namita Tare, Prachiti Vartak, Kimaya Rakvi,
Department of Computer Engineering
Thakur College of Engineering & Technology
Mumbai, India
vikas.singh@thakureducation.org, tarenamita3@gmail.com, vartakprachiti84@gmail.com, kimrakvi@gmail.com

Abstract—**The healthcare industry collects a big amount of data which is not properly mined and not put to the optimum use. Discovery of these all hidden patterns and relationships often goes unexploited. Our research focuses on the all aspect of Medical diagnosis by learning pattern through the collected data of heart diseases and to develop intelligent medical decision support systems to assist the physicians. Cardiovascular disease remains the biggest cause of deaths worldwide and the Heart Disease Prediction at the early stage is important. Heart disease is one of the deadly diseases. A large population in the world is suffering from this disease problem. As we consider the death rate and a large number of people who are suffering from heart disease, it is disclosed how important is early diagnosis of heart disease problem. There are many traditional methods of prediction for such illness but they are not looking sufficient for every people. There is an urgent need of medical diagnosis system that can predict the heart diagnosis problem at an early stage and offers more accurate and correct diagnosis result than traditionalmethods.**

***Keywords*— heart disease, Convolution neural network, Decision tree, Random Forest, DeepLearning.**

I. INTRODUCTION

The healthcare industry collects a huge amount of data which is not properly mined and not put to the optimum use. Discovery of all hidden patterns and relationships often goes unexploited. This situation can be overcome by Advanced data mining modeling techniques .There is a huge amount of untapped data that can be turned into useful important information. The decision support systems that have been developed to assist physicians in the diagnostic process often are based on static data which may be out of date. A decision support system which can learn the relationships between history of patient, diseases in the population, symptoms, pathology of a disease, family history and test results, would be helpful to physicians and hospitals. The concept of Decision Support System (DSS) is very broad because of many different approaches and a wide range of domains in which decisions are made. DSS terminology refers to a class of computer-based information systems which will include knowledge-based systems thatsupportdecision making activities. In general, it can say that a DSS is a computerized system for assisting to make decisions. A DSS application can be composed of subsystems. However, the development of such type of system presents a daunting and yet to be explored task. Many factors have been attributed but insufficient information has been identified as a major challenge. To reduce the diagnosis time and improve the diagnosis accuracy, it has become more of a demanding issue to develop reliable and powerful medical decision support systems (MDSS) to provide support the yet and still increasingly complicated diagnosis decision process. The medical diagnosis by nature is a complex and fuzzy cognitive process, therefore soft computing methods, such as decision tree classifiers has shown great potential to be applied in the development of MDSS of heart diseases and other diseases. The main aim is to identify the most important risk factors based on the classification rules to be extracted.

II. REVIEW OF LITERATURE

Amin et al. Proposed a hybrid system in which genetic algorithm and ANN is used. This system predicts the heart disease problem based on the risk factor. They trained ANN in a similar way i.e. Using the back propagation as trained previously. They mainly highlight that two major disadvantages of the back propagation algorithm. First one is that to finding out the initial weights which are globally optimized is almost impossible. Second, in convergence back propagation is slow. To solve these type of issues, they applied the genetic algorithm to optimize the weights of ANN and got the better performance than simple ANN. They successfully achieved an accuracy of 96.2%

on the training set and the accuracy of 89% on the testingset.

Sarath Babu, Vivek EM et al. Proposed Heart Disease Diagnosis System Using Data mining technique. Data mining is an advanced technology, which is the process of discovering actionable main information from large set of data, which is used to analyze the large volumes of data and extracts patterns that can be converted to the useful knowledge. Medical data mining has a great potential for exploring all hidden patterns in the data sets of medical domain. These all the patterns can be utilized to do clinicaldiagnosis. All the data need to be collected in a standardized form. From the medical profiles fourteen attributes are extracted such as age, sex, blood pressure and blood sugar, obesity etc. Can predict the likelihood of the patient getting heart disease. These attributes are fed in to the K-means algorithms, MAFIA algorithm and Decision tree classification in heart disease prediction, applying the data mining technique to heart disease treatment; it can provide as reliable performance as that achieved in diagnosing heart disease problem.

Theresa Princy. R , J. Thomas proposed a system for Human Heart Disease Prediction using Data Mining Techniques. The main motivation of this paper is that to provide an insight about detecting heart disease risk rate using data mining techniques. Various Data mining techniques and classifiers are discussed in many studies which are used for the efficient and efficacious heart disease diagnosis. As per the analysis mode, it is seen that mostly many authors use various technologies and different number of attributes for their study. Hence, All different technologies give different precision depending on a number of attributes considered. Using KNN and ID3 algorithm the risk rate of heart disease problem was detected and accuracy level also provided for different number of theattributes.

Jayshril S. Sonawane , D. R. Patii proposed a system which Predict Heart Disease Using Learning Vector Quantization Algorithm. In this paper an efficient system for heart disease prediction is developed. This system accepts the 13 clinical features as input and the training of network is performed by using the Learning Vector Quantization algorithm. It gives the presence or the absence of heart disease. The system also computes various different parameters like accuracy, sensitivity, specificity, training time, testing time, error and ROC curve. To improve the performance of the heart disease prediction system, it is trained with different number of neurons and also the number of training epochs are varied. The results shows that our system when we compared with others it gives the highest accuracy of 85.55%.

III. RESEARCH ALGORITHM

A. Deep Learning

Deep learning is also known as deep structured learning or hierarchical learning. It is a part of a broader family of machine learning methods based on artificial neural networks. Learning can be in the form of supervised, semi-supervised or unsupervised. Deep learning architectures such as deep neural networks, deep beliefnetworks, the recurrent neural networks and convolutionalneural networks have been applied to various fields which include computer vision, speech recognition, naturallanguage processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and boardgame programs, where they have produced results comparable to and in some cases that are superior to human experts.*Convolution NeuralNetwork*CNN is the class of deep and feed-forward artificial neural networks where connections between nodes do not form a cycle. CNNs are mostly used in computer vision, however they've shown great results when applied to various NLP tasks as well. CNN's are performing good at extracting local and position-invariant features. CNN's work well for tasks where sequential modelling is more important. Choose a CNN for classification tasks such as sentiment classification since sentiment is generally determined by some key phrases .The structure of CNN includes two layers one is feature extraction layer and second is feature map layer. In feature extraction layer, the input of each neuron is connected to the local receptive fields of the previous layer, and extracts the local feature. Once the local features are extracted, the positional relationship between it and other features also will be determined. Every feature map is a plane where the weight of the neurons in the plane are identical. The neurons in the same feature map plane have the equal weight, so that the network can studyconcurrently.

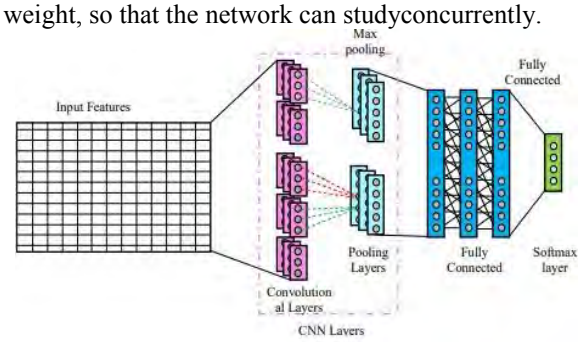


Figure 1: The architecture of CNN

B. Decision Tree

A decision tree is a flow-chart similar to tree structure, where each internal node represents a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions. The topmost node in the tree is called as root node. Internal nodes are represented by rectangles, and leaf nodes are represented by ovals. For the classification of an unknown sample, the sample attribute values are being tested across the decision tree. A path is drawn from the root to a leaf node which holds the class prediction for that

sample, so a decision trees can easily be converted to classificationrules.

C. RandomForest

Random forest is a decision tree which is a decision support tool. It uses a tree like model of decisions It is predominantly an band of unprimed classification trees. That provides remarkable performance on a number of practical problems Such as health care predictionproblems.it is not sensitive to noise ,in data set. It is not being subjected to over fitting. It is built by combining predictions of several trees. Each of them are trained separately. It works fast. Usually exhibits a great performance improvement over many other tree-based algorithms such as a decision tree. Three main choices to be made a random treeare

- Splitting theleaves.
- Predictor is to use in eachleaf.
- Inject randomness into the trees.KNN

IV. PROPOSED MODEL

Heart Disease prediction system is developed using the convolution neural networks. The main benefit of CNN is that CNN has high discriminative power. The flowchart of the proposed model is shown below in figure 2.

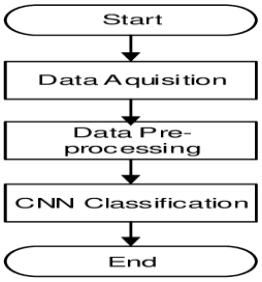


Figure 2: Flowchart of heart disease prediction based on CNN

A. Dataset

Data Set used for proposed system is obtained from UCI (University of California, Irvine C.A) Centre for machine learning and intelligent systems

Sr No	Risk Factors	Values
1	Age	Age of patient in year
2	Sex	Value=1: Male value=0:Female
3	Chest pain type	Value=1:typical type 1 angina Value=2: typical type angina Value=3: non-angina pain Value=4:asymptomatic
4	Resting blood pressure	In mm Hg on admission to the hospital
5	Serum Cholesterol	mg/dl
6	Fasting blood sugar	Value=1: > 120mg/dl Value=0:< 120mg/dl
7	Resting Electrocardiographic results	Values=0:normal Value=1: 1 having ST-T wave abnormality Value=2:showing probable or definite left ventricular hypertrophy
8	Maximum heart rate achieved	Heart rate
9	Exercise induced angina	Value=1:yes value=0:no
10	Old peak	ST depression induced by exercise relative to rest
11	The slope of the peak exercise ST segment	Value=1: unsloping value= 2: flat value=3:down sloping
12	Number of major vessels colored by fluoroscopy	Value in range of 0-3
13	Thal	value3 = normal value 6 = fixed defect value 7 = reversible defect

Table 1: Attributes of heart disease risk prediction.*Performance EvaluationCriteria*

Four different metrics i.e. accuracy, precision, recall, and F1-measures are used to evaluate performance of the proposed system. We denote TP, FP, TN, and FN as true positive means the number of instances correctly predicted as required, false positive means the number of instances incorrectly predicted as required, true negative means the number of instances correctly predicted as not required and, false negative means the number of instances incorrectly predicted as not required respectively. Then, we can calculate four metrics: accuracy, precision, recall, and F1- measure asfollows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\textbf{Precision} = \frac{TP}{TP + FP}$$

$$\textbf{Recall} = \frac{TP}{TP + FN}$$

$$\textbf{F - measure} = \frac{2*Recall*Precision}{Recall + Precision}$$

Accuracy, Pricision, Recall and F1-Measure are used to convey the success of pedicting Heart Disease. Using only Accuracy can be sometimes misleading. Sometimes selecting model which has lower accuracy is desirable, becouse it provide more robust predictor for the problem. All predictions can be predicted as the value of majority class by model when the problem domain has a huge class imbalance. Therefore, we are going to prefer the four different factor to get more accurateresults.

B. ExperimentalResults

Data Set used for proposed system is obtained from UCI Centre for machine learning and intelligent systems. Data Cleaning as well as filtering is done to remove duplicate records, normalize the values, accounting for missing data and removing irrelevant data items. So here we implement Decision Trees, Random Forest and Convolution Neural Network algorithm in our project. Our project will predict heart disease based on Convolution Neural Network but As well as these we also implement Decision Tree and Random Forest algorithm for comparing accuracy of the algorithm.

The Algorithm Use	Accuracy (Training Set)	Accuracy (Testing Set)
Random forest	98.58 %	80.2 1%
Decision Tree	100 %	72.5 2%

Figure 2: Accuracy of Random Forest & Decision Tree

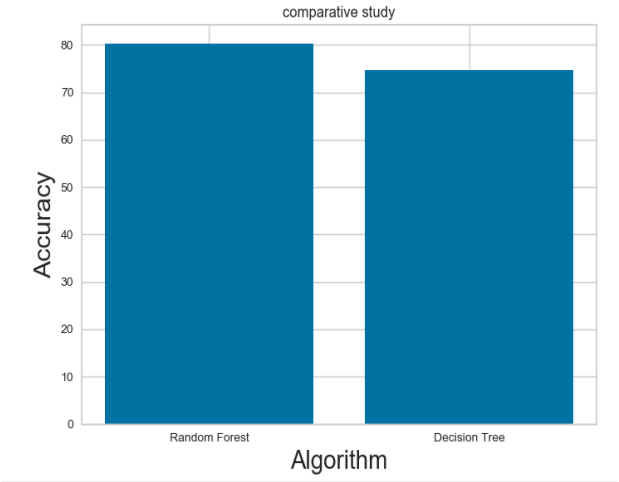


Figure 3: Comparative study of DT & RF

V. CONCLUSION

We have proposed Supervised Learning Algorithm for finding the risk of heart disease of a patient using the information collected from the patients. CNN has a different feature that it does not use a predefined set of hidden units, instead the hidden units gets added up one by one until the error is minimized. By exploiting this distinct feature of the CNN, a computerized prediction algorithm is developed that are not only accurate but also computationally effective for heart attack prediction with the proper adaptation of CNN classifies, the method can thus evolve number of hidden units within an architecture space. The neural network approach to generate accurate classification rules is proposed. To perform classification task of medical data, the neural network is trained using Convolutions Neural Network algorithm. The experiment is conducted with heart disease dataset of UCI by considering the single and multilayer neural network modes. Convolution neural network algorithm is the multilayer perceptron that is thespecial design for identification of two-dimensional image information. Always has more layers that ares input layer, convolution layer, sample layer and output layer. In addition, in a deep learning network architecture the convolution layer and sample layer can havemultiple.

References

[1] Syed Umar Amin , Kavita Agarwal , Dr. Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors ", Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT2013)

[2] Sarath Babu, Vivek EM et al, Heart Disease Diagnosis Using Data Mining Technique ", International Conference on Electronics, Communication and Aerospace Technology ICECA2017

[3] Theresa Princy. R , J. Thomas," Human Heart Disease Prediction System using Data Mining Techniques", 2016 International Conference on Circuit, Power and Computing Technologies[ICCPCT]

[4Jayshril S. Sonawane , D. R. PatiI, "Prediction of Heart Disease Using Learning Vector Quantization Algorithm ", 2014 IEEE

[5] SP Rajamahona,C. Akalya Devi et al, " Analysis of Neural Network Based Heart Disease Prediction System",2018 IEEE

[6] Sana Bharti, Dr.Shaliendra Narayan Singh, "Analytical Study of Heart Disease Prediction Comparing With Different Algorithms "International Conference on Computing, Communication and Automation(ICCCA2015)

[7] Mamatha Alex P, Shaicy P Shaji, " Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique ", International Conference on Communication and Signal Processing, April 4-6, 2019,India

Arduino Based Smart Blind Stick

Sandeep Pandey, Prem Patel, Riddhi Patel
Information Technology, Thakur College Of Engineering and Technology, Mumbai, India
Sp348220@gmail.com, patelprem616@gmail.com, patelriddhi1519@gmail.com

Abstract— **In the present scenario, a visually impaired individual is constrained to depend on someone else and can't make a trip autonomously to wherever without the assistance of others. So, here we propose ansmart blind stick that allows visually challenged people to navigate with ease using advanced technology. The blind stick is integrated with ultrasonic sensor, soil moisture sensor along with tilt sensor. On detecting obstacle, the sensor passes this information to the microcontroller. The microcontroller then processes this information and figures if it is sufficiently close If the obstacle is close the microcontroller sends a signal to sound a buzzer. If the obstacle isn't that nearby the circuit sits idle. It also detects water or mud and sounds a buzzer in different pattern and alerts the blind. By using tilt sensor blind man could remotely locate his stick. We have also used a motor vibrator which provides different vibration pattern for alerting the user.Thus, proposed system helps the people who are visually impaired in the aspect of mobility and making it safer by avoiding accidents.**

Keywords – Ultrasonic sensor, Soil Moisture Sensor, Tilt Sensor, Arduino UNO, Smart Blind Stick

I. INTRODUCTION

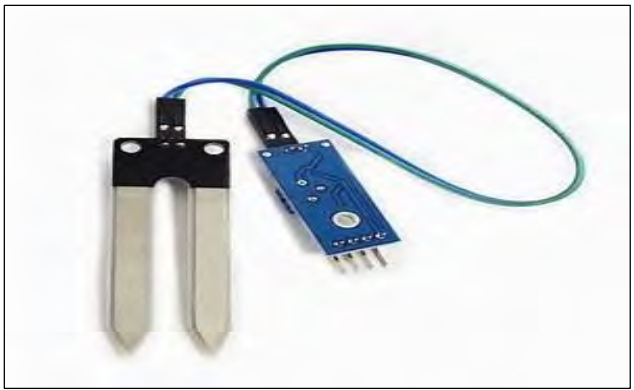
Visual deficiency is a common regular inability among the peoples throughout the world. As per the World HealthOrganization (WHO) 285 million individuals are outwardly impeded around the world, 39 million are visually impaired and 246 have low vision. About 90% of the world's outwardly impeded live in developing nations.Visually impaired persons have difficulty to interact with their environment. They have a very little contact with surroundings. Physical movement or mobility is a challenge for visually impaired persons, because it can become difficult to recognize obstacles appearing in front of them, and they are not able to move from one place to other. Smart Blind Stick is specially designed to detect obstacles which may help the blind to navigate care-free. The audio messages will keep the user alert and comparatively reduce accidents. The proposed system contains the ultrasonic sensor, Soil moisture sensor, Buzzerand Arduino UNO. The main objective of the paper is to discuss about development work of a stick that could communicate with the users through voice alert and vibration, which is named Smart Blind Stick which involves coding as software and physical installation as hardware part.

II. PROBLEM DEFINTION

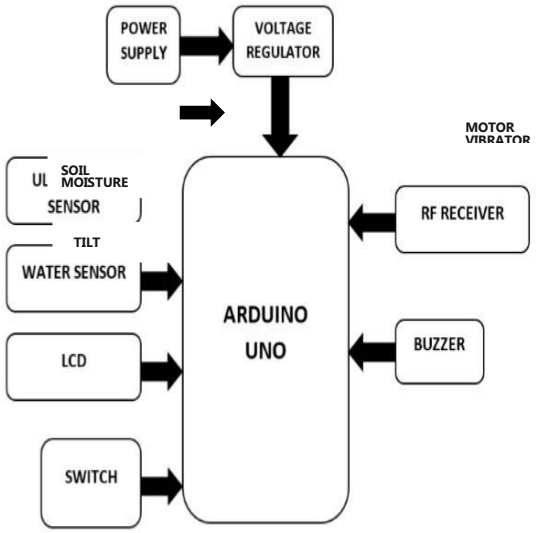
Physical movement is a challenge for visually impaired person. The conventional walking stick used by them is inefficient in detecting obstacle. The disadvantage of conventional cane,However,is its failure to detect obstacles outside of its reach. That is the user has to tap to the ground or obstacle to detect it. The visually challenged people can avoid the object better if the walking stick can produce sound and vibration warnings when there is a object in specified range.

III. BACKGROUND

More than 161 million people worldwide are visually impaired. Among them, 124 million have Low vision and 37 million are blind. Another 153 million people suffer from visual impairment Due to uncorrected refractive errors such as near-sightedness, far-sightedness or astigmatism. Virtually all these people could restore normal vision with eyeglasses or contact lenses. More than 90% of the world's visually impaired people live in low- and middle-income countries. Blindness is a condition of lacking visual perception and it is always described as severe visual impairment with residual vision. The blind people's life and activities are greatly restricted by loss of eyesight. They can only walk in fixed routes that are significant in their lives, with blind navigation equipment's and the accumulated memories in their long-term exploration. This situation has resulted in many difficulties to them in their normal work, lives, activities, and so on. Based on the investigation about daily activity characteristics and modes of the blind, the study found that the main difficulties encountered in a trip of the blind included walking on the road, finding way, taking a bus and looking for usual life-arena Several devices have been developed for mobility and navigation assistance of the blind and are typically known as travel aids or blind mobility aids.

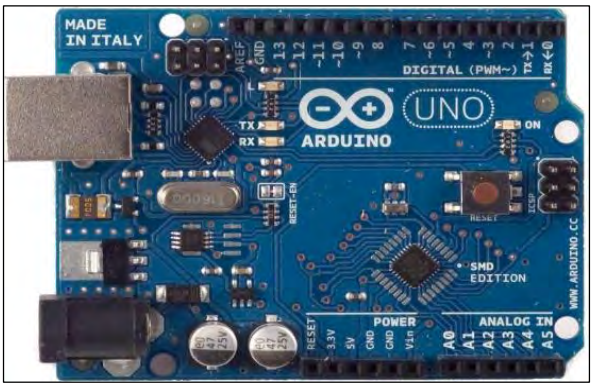


IV. BLOCK DIAGRAM

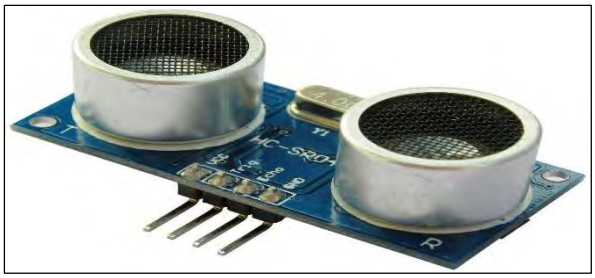


V. HARDWARE REQUIREMENTS

1. Arduino UNO:



Arduino UNO is a microcontroller board based on ATmega328p.It has 20 pins out of which 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz quartz crystal, a USB connection, a power jack, an ICSP header and a reset button. It contains everything needed to support the microcontroller. The Operation Voltage is 5V simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started. You can tinker with your UNO without worrying too much about doing something wrong.



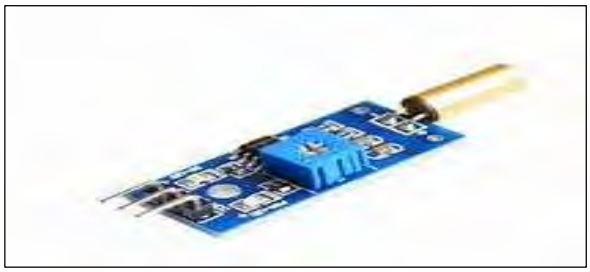
2. Ultrasonic Sensor:

As the name suggest, ultrasonic sensors measure distance by ultrasonic waves. The sensor head emits an ultrasonic wave and receives the wave reflected back from the targeted object. Ultrasonic Sensors measure the distance to the object by measuring the time between the emission and reception. An optical sensor has a transmitter and receiver, whereas an ultrasonic sensor uses a single ultrasonic element for both emission as well as reception. In a reflective model of ultrasonic sensor, a single oscillator emits as well as receives ultrasonic waves alternately.

3. Soil Moisture Sensor:

Soil moisture sensors measure the volumetric water content in soil. The direct measurement of free soil moisture process includes drying, removing, and weighing of a sample.Soil moisture sensors calculates the volumetric water content indirectly by using various property of the soil, such as dielectric constant ,electrical resistance or interaction with neutrons, as a proxy for the moisture content.Soil moisture sensors is used to estimate or calculates volumetric water content. There is a separate class of sensors that measure another property in soils called water potential; these are usually known as soil water potential sensors and which includes gypsum blocks and tensiometers.

4.Tilt Sensor:



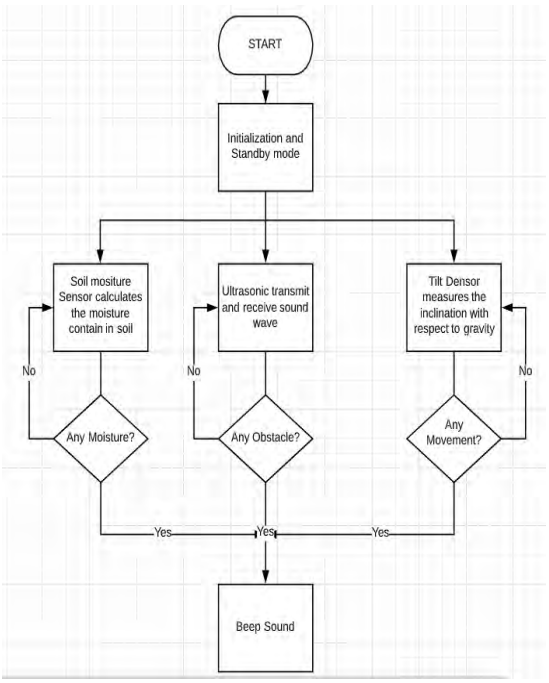
A tilt sensor is an instrument that is used for measuring the tilt in multiple axes with respect to particular plane. Tilt sensors measure the tilting position with respected to gravity and are used in number applications. They enable the easy detection of orientation or inclination of a object. Similar to mercury switches, they also be known as tilt switches or rolling ball sensors. The capability of tilt sensors is influenced by factors together with vibration, gravity, linearity, 0 offset, temperature, cross-axis sensitivity, acceleration/deceleration, shock, clean line of sight between the user and the measured point, and calibration of tilt sensors.

5.Buzzer:

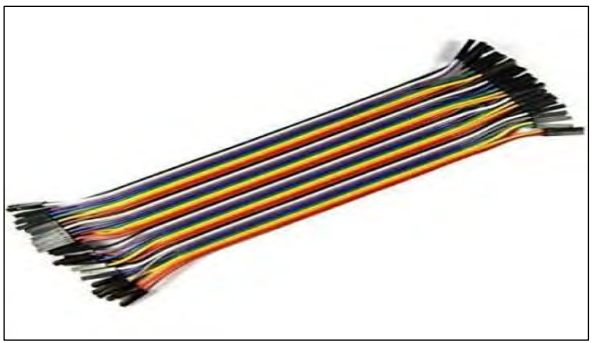
A buzzer is a small efficient component to add sound to our project/system. It iscompact and smallwith 2-pin structure hence can be easily mounted on breadboard, Perf Board and even on PCBs which makes this a widely used component in most electronic applications. There are commonly two types of buzzers available. First one is a simple buzzer which will make a Continuous Beeeeppp.... sound and other type of buzzer is called as readymade buzzer which looks bulkier than simple buzzer and produce a Beep. Beep. Beep. Beep Sound is produced because of the internal oscillating circuit present inside it.

6.Bread Board:

A breadboard is a solderless device for prototype model



with hardware and test circuit structures. Components in electronic circuits can be interconnected by inserting their terminals into the holes and then making required



connections through jumper wires. The breadboard has metal strips underneath the board and which connect the holes on the top of the board.

7. Jumper Wires:

A jumper wire is a conducting wire which is used to transfer electrical signals between two ends in a circuit. The wires can be used to modify circuits as well as to diagnose problems within a circuit. Jumper wires typically vary in size and color depending on what they are being used for. Jumper wires are used to establish connections between the central micro controller and other devices such as sensors and buttons..

VI. SOFTWARE REQUIRMENTS

The Arduino Integrated Development Environment (IDE)is a cross-platform application for Windows, Linux, macOS that is written in functions from C and C++. It is used to write and upload various programs to Arduino compatible boards, also, with the help of third party cores, other vendor developmental boards.

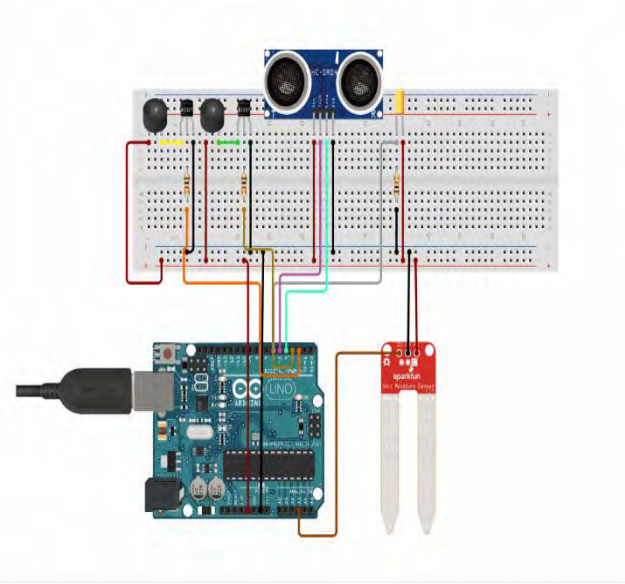
The source code for IDE is released under the GNU General Public License v2. The Arduino IDE supports the languages C++ and C with special rules of code structuring. The Arduino IDE supplies a software library from the Wiring project, which provides many input-output procedures. User defined code requires two basic functions, for starting the sketch and the main program loop, that are compiled and linked with a program into an executable cyclic executive program with the GNU tool-chain, also included with the distribution of IDE. The Arduino IDE provides the program avrdude to convert the executable code into text file in hexadecimal encoding format that is loaded into the Arduino board by a loader program in the firmware of board.

VII. DESIGN AND DEVELOPMENT:

1. Flowchart:

When circuit is switched ON the sensors placed on the stick goes intoactive state. There are three sensors mounted on the stick which are ultrasonic sensor to detect obstaclein front of the stick, soil moisture sensor to detect water on the ground and tilt sensor which enables the user to locate the stick from ground. Ultrasonic sensor transmit and receive sound wave which in turn helps to detect obstacle. If any obstacle is detected within a specified range then it will produce a beep sound or else do nothing. Soil moisture sensor calculates moisture contain in soil.is any moisture contain is detected in the soil then the stick will produce beep sound or else do nothing. Tilt sensor measures the inclination of the stick with respect to gravity. If the stick is inclined then it will produce a beep sound or else do nothing.

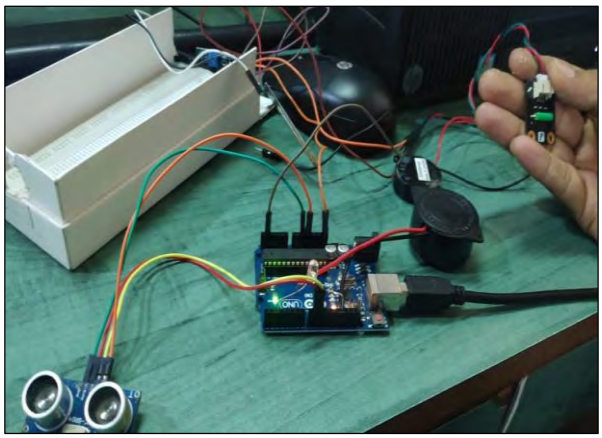
2.Circuit Diagram:



VIII. WORKING OF SYSTEM

Smart Blind Stick is basically designed to make sure that the people with visual disabilities can work on their

own without being dependent on others for their living. This innovative Stick allows visually challenged people to navigate on their own. This blind stick is integrated with Ultrasonic Sensor so that people with disabilities can walk with a sense of obstacle in between. This ultrasonic sensor makes sure that when there is any obstacle in front of the stick it indicates the user. Ultrasonic sensor sensesthe obstacle and passes this data to microcontroller. Then this microcontroller will process the data and calculate if the obstacle is close enough or not. If the obstacle is not close enough then it does nothing. If obstacle is close enough then it will produce a buzzing sound so that the user can understand that there is obstacle in front. Buzzer is used so that it can produce a sound when obstacle is ahead. This stick also contains soil moisture sensor. Soil moisture sensor is used so that the user can indicate the path where there is no detection of water and hence



selecting a simple path to walk. Soil moisture sensor is connected at the bottom of the stick and when the sensor detects water in the soil then it produces a buzzing sound to indicate the user that the path is not safe. We made use of vibrator motor for the people with hearing as well as visual disabilities. This motor vibrates while buzzing which in turn helps them to detect obstacle in between. This stick includes a tilt sensor which will help to locate the stick on the ground. Tiltsensor is of a great use to the people with visual disabilities. When this stick falls of the ground from the hand of the user then this stick produces a buzzing sound so that the user can easily locate the stick. This tilt Senor measures the angle of the stick with respect to gravity and then detect the inclination to decide whether to produce buzzing sound or not. This stick also contains a LDR which helps to produce light on the stick at times of dark which will allow the people surrounding the user to be aware and hence avoid collation.

This project is designed in such a way that the people with visual disabilities can easily use this stick and be independent. The look of this project is just like a normal stick with some sensors attached which does not give any problem to the user to use this stick.

IX. FUTURE SCOPE

By adding a GPS and GSM module to the stick we can track the actual location of the person as well as establish communication between mobile device or a computing machine. We can also add Vibrator motor which helps to make a vibration by recognizing it the person holding the stick and also feel the vibration and act accordingly, by this motor this stick can be used by people with hearing disabilities.

X. CONCLUSION

The project proposed alayout,design and architecture of a new concept of Smart Electronic Guiding Stick for blind people. It aims to solve the problems faced by the millions of visually impaired people worldwide in their daily life. The proposed system takes measures to ensure their safety and the main advantage of the system is that it can prove to be very low-cost solution to millions of visually impaired people worldwide with the proposed combination of various working units. It is a real-time system that monitors position of the user and provides dual feedback making navigation more safe and secure than the traditional cane. It can be further improved to have more decision taking capabilities by integrating various types of sensors such as GPS, GSM module and thus could be used for different applications.This our hope is to consider this stick as smart eye for the visual impairments. It is worth mentioning at this point that the aim of this study which is the design and implementation of a smart walking stick for the blind has been successfully achieved.

References

1.https://circuitdigest.com/microcontroller-projects/arduino-smart-blind-stick

2.https://www.instructables.com/id/Smart-Blind-Stick/

3.https://electronicsforu.com/electronics-projects/smart-stick-using-arduino-ultrasonic

4.https://create.arduino.cc/projecthub/hadi1234/arduino-blind-stick-19d865

5. Dambhare, Shruti, and A. Sakhare. "Smart stick for Blind: Obstacle Detection, Artificial vision and Realtime assistance via GPS." In 2nd National Conference on Information and Communication Technology (NCICT).2011; 2:31-33.

6. Al-Shehabi, Monther M., Mustahsan Mir, Abdullah M. Ali, and Ahmed M. Ali. "An Obstacle Detection and Guidance System for Mobility of Visually Impaired in Unfamiliar Indoor Environments." International Journal of Computer and Electrical Engineering. 2014;6(4): 337-341.

7. Jeslin, J. Gnana, A. Vaishnavi, and J. Nivedha. "Enhancement of Smart Cane To Detect And Notify Obstacles.". International Journal of Engineering Research & Technology (IJERT) 2015;4(2):1-3.

Text Extraction Based On Optical Character Recognition And Natural Language Processing

Prateek Angadi , Devavrat Singh Bisht , Khushi Dubey , Esha Ghorpade , Anupriya Babbar Assistant Professor
Thakur College Of Engineering and Technology Mumbai-400101 INDIA
prateekangadi66 @gmail.com , devavratsinghbish t@gmail.com , khushidubey2k@ gmail.com , eshag06@gmail.com , anupriyababbar.86@gmail.com

1. Abstract: OCR sands for Optical Character Recognition, it is the automated translation of images, printed or handwritten text into string, whether from a scanned document, a photo of a document or from subtitle text overlying on an image. In OCR first the text regions which are present in camera captured image or in the scanned copy are extracted and skew corrected. Then the regions with text are binaries and divided into parts, lines and characters. By using OCR system we can retrieve text from camera captured image or from handheld scanner andetc.

Natural language processing (NLP) is widely known for representing and analyzing human language by automatic manipulation of natural language, like speech and text, by software. It has a varied range of applications such as in fields like machine language translationwhichinvolvestranslatingdifferentforeign languagesintotheonerequiredbytheuserwhicheases communication, detecting spam emails, information (processed data) extraction and summarization, medical purposes etc.
The paper discusses about the different levels of NLP and components of Natural Language Generation (NLG) followed by discussing about how NLP can be used with OCR to enrich the process of intelligent document imaging by aiding in recognizing relevant conceptsintheinputtextwhichisbeneficialinsystems.

Using OCR and NLP text extraction can become very easy as OCR extracts the whole text and NLP is used to extract specific text from them.

Keywords: Optical character Recognition (OCR), Natural Language Processing (NLP), Natural Language Generation (NLG).

2. INTRODUCTION:

2.1 OCR:

A few decades ago, research in OCR was only limited to the document images acquired from Flat-bed scanners, but now we can use this system in every person's phone as every person's phones are acquainted with cameras and are portable. Due to the advancement of processing speed and internal memory of smartphones having digital camera give real time results.
But implementing OCR in hand held devices is not an easy task as images acquired from these digital cameras often suffer from skew and perspective distortion.
In addition to that, not everyone is professional in using this system, so while they perform capturing process, uneven and insufficient illumination and improper focusing yield poor quality images.

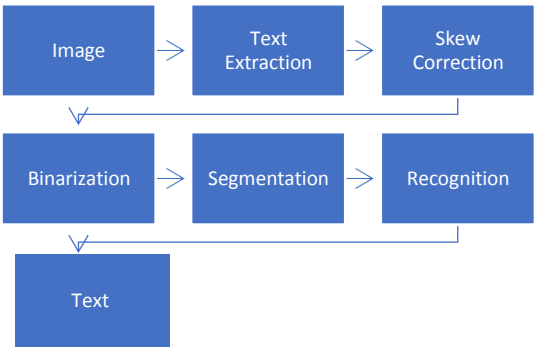


Fig 1. Process inside OCR

2.2 Natural Language Generation

Natural-language generation (NLG) is a softwareprocessthattransformsstructureddata or information into natural language which involvesconvertinginformationfromcomputer databases or semantic intents into language readable by humans. Its main use can be in producing long form content for organizations which helps in automating custom reports, as well as produce custom content for a web or mobile application. It can also be used to initiate short interactive conversations (a Chabot) with a human friendly machine which might even be read out by a text-to-speechsystem.

3 PRESENTWORK:

Presently the health reports are verified and data is entered into the database manually which consumes a lot of time and man power. Eachandeverykeyvaluepairisenteredstepby step by anemployee.

4 OURIDEA:

This process of manually entering of data into database can be done using by using two technologies simultaneously, that is OCR for recognition of characters from a set of images andNLPforextractingkeyvaluepairsformthe string provided by OCR. These key value pairs willbethenfeedintothedatabase.Throughthis method the whole process becomes faster affecting the waiting time of the clients of insurance companies and thus the overall thorough-put of thecompany.

4.1 Working ofOCR:

Image is nothing but a matrix of pixels, each pixel can have three different parameter on which they glow up, the parameters being the intensity of RGB (Red Green Yellow) colors with which they glow these values range from 0 to 255.

These values that range from 0 to 255 are first divided by 255 so that they come in a range of 0 to 1, this makes it easy to train the neural network. The NN is then trains over a set of dataset and makes predictions. The quality of NN thus trained is measured by it accuracy while predicting on a data that is similar to nut notexactlythetrainingdataset,thesedatasetare known as test dataset. The train dataset andtest dataset usually has a ratio of7:3.

Also while training a NN over a set of image the resolution of the image plays a very Importantrole.Thenoofcolumnsinthematrix mustbeequaltothenumberofnodesatthefirst layer of the NN. There are many hidden layers between the first and the last layer and at all layer a prediction takesplace.

OCR uses these previously trained NN tomake predictions and thus turn pictures intostring.

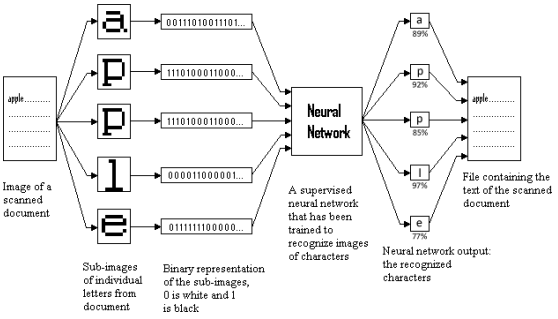


Fig 2. Training of Neural Network

Feature Extraction: In this stage, various features of characters are extracted. These features uniquely identify characters. The selection of the right features and the total number of features to be used is an important research question. Different types of features such as the image itself, geometrical features (loops, strokes) and statistical feature (moments) can be used. Finally, various techniques such as principal component analysis can be used to reduce the dimensionality of the image.

APPLICATIONS OF OCR:

OCR enables a large number of useful applications. Duringthe earlydays,OCRhasbeenusedformailsorting, bank cheese reading and signature verification Besides, OCR can be used by organizations for automated form processing in places where a huge number of data is available in printed form. Other uses of OCR include processing utility bills, passport validation, pen computing and automated number plate recognition. Another useful application of OCR is helping blind and visually impaired people to readtext.

4.2 NLP:

NLP in 4 Waves.

The history of natural language processing(NLP) can be timed back to 1950's where Alan Turing published an article titled "Computing Machinery and Intelligence" which proposed what is now called the Turing test as a criterion of intelligence. Mostnatural

Language processing systems were based on complex sets of hand-written rules in the past and the actual revolution in NLP was brought about in the late 1980's when machinelearning algorithms began to gain prominence for processing information. Scientists Richard Bundler and John Grinder started this as the basis of NLP around 1972 with modellingFritz Perl's (Gestalt Therapy), Virginia Stair (Systemic Family Therapy) and Milton Erickson(Hypnotherapy).

APPROACHES: The various approaches to natural language processing can be categorized into

1) Distributional 2) Frame-based 3) Model-

theoretical

4) Interactive learning

4.2.1 DISTRIBUTIONAL APPROACHES

Large-scale statistical tactics of machine learning and deep learning are included in distributional approaches. These methods basically converts content or input words into word vectors for mathematical analysis and performextremelywellattaskssuchaspart-of- speech tagging (identifying the word category inEnglishgrammar),dependencyparsing(does this part of a sentence modify another part?), and semantic relatedness (are these different words used in similar ways?). These NLPtasks don't depend on understanding the meaning of words, but relation betweenwords.

Such systems are broad, flexible, and scalable. They can be applied to most of the documents presentinthemarketwithoutexpertknowledge

In the domain.

4.2.1 FRAME-BASEDAPPROACH

Framesareacanonicalrepresentationforwhich specifics that can beinterchanged.

Sentences that are syntactically different but semanticallyidenticali.e.theyconveythesame meaning.

4.2.1 MODEL-THEORETICAL APPROACH

Model-theoretical approach is the third category of semantic analysis. "Model theory" and "compositionality" are the two important linguistic concepts to understand thisapproach. Model theory refers to the idea that sentences refer to the world, for example with grounded language (i.e. the block is blue). Meanings of the parts of a sentence can be combinedto deduce the whole meaning, this is knownasCompositionality.

4.2.1 INTERACTIVE LEARNING

British philosopher of language (Paul Grice), described language as a cooperative game between speaker and listener. It is believedthat a viable approach to tackling languagelearning istoemploydynamic,interactiveenvironments where humans teach computers gradually. In such approaches, the pragmatic (purpose and goals) needs of

Development.

- Rule-based.
- Traditional Machine Learning.
- Neural Networks.
- Text ClassificationTasks.
- Word SequenceTasks.
- Text MeaningTasks.
- Sequence to SequenceTasks.
- DialogSystems.

5. USES OF NATURAL LANGUAGE PROCESSING WITHOCR

In our project we aim to digitize different kind s of printed lab reports / health check reports to discrete values of each parameter captured. Conversion of data to a dataset of digitized value will help in automating lot of back-end approval process. Ability to read and classify comments/observations for common parameters will be implemented using natural language processing. OCR is used to translate printed, hand written, and scanned documents into a machine-readable format. The technology relieves employees of manual entry of data, cuts related errors, and enables automated data capture and we are going to use the same technology in digitization of lab reports by identifying key value pairs for the data entered. OCR is the core technology for automatic text

recognition. The technology relieves employees of manual entry of data, reduces errors, and provides automated data capture. Information in documents is usual y a combination of natural language and semi- structured data in forms of tables, diagrams, symbols etc. with NLP machines are able to inter act with the written form of human language. NLP along with the OCR technology, finds apeplacations for data retrieval, information extraction, and text summarization. For training a do cement classifying model, it is necessary to have thousands of samples that will make a training dataset. Then, machine learning engineers defines the parameters according to which it is possible to associate a document with one or another class and use these parameters later to create a classification model. Different training methods on deep neural nets can make a good fit. Everything depends on the type ofdata.

So in our case we will be using Tesseract OCR by Google demonstrates outstanding results enhancing and recognizing raw images, categorizing, and storing data in a single database for further uses. It supports more than 100 languages, and the accuracy of document recognition is good enough for somecases.

6. COMPARATIVE STUDY:-

This is a new technology. Here thephotoofthelab report will be uploaded and the text in the image will be extracted. The key values of the text will be extractedand Stored in the database of the company. All this work willbe done automatically and the time for doing this work willbe Reduced tremendously.	In the recent tech all the updates are to be done manually which is really very hectic and time Consuming.
The verification of the data is done twice. The data is Secured.	There is no verification for the data.

7 CONCLUSION:

In this paper, an overview OCR and NLP and howtheycanbeusedtogetherfordataentryhas been given. Different parts of OCR and NLP work together for this to happen. Also the accuracy of this project depends a lot on the ability of OCR to recognize text correctly and the power of NLP to extract the key valuepairs accurately.

Human Activity Recognition Smart watch

Prof. Prerna Deshmukh Riya Raipure Gayatri Wandhare,
Nirbhay Dawle Akshita Sagole Jay Darwankar
Priyadershani technology Collage of Engineering,
Nagpur

Abstract: **Sensor-based activity recognition seeks the profound high-level knowledge about human activities from multitudes of low-level sensor readings. Conventional pattern recognition approaches have made tremendous progress in the past years. Recently, the recent advancement in technology makes it possible to perform automatic high-level feature extraction thus achieves promising performance in many areas. Through a smart watch we tend to monitor human activities using different aspects. We also present detailed insights on existing work and propose grand challenges for future research. Automatic high-level feature extraction thus achieves promising performance in many areas. Through a smart watch we tend to monitor human activities using different aspects. We also present detailed insights on existing work and propose grand challenges for future research**

Keywords: Step tracker, Sensors, HAR, and Pulse.

I. INTRODUCTION:

Human Activity Recognition (HAR) is the event of a device where a particular device uses various types of sensors to detect human activity movements, collects the data, analysis it and then gives a brief result.

In today's time there are various devices that collects the data and predicts the Outcome with a ratio of probability and then displays result.

Its uses various aspects of deep learning and some parts of neural networks to connect the dataset which are used.

The data sets are the background activity where the actual work of recognition and results are given. We propose to build an application that will help to monitor different human activities to insure the health concerns.

The idea is to make an application for a low cost affordable smart watch which can count the number of steps on basis of your walk and then store the data and then monitor the pulse rate for the particular person and then monitors and verify the health of the person.

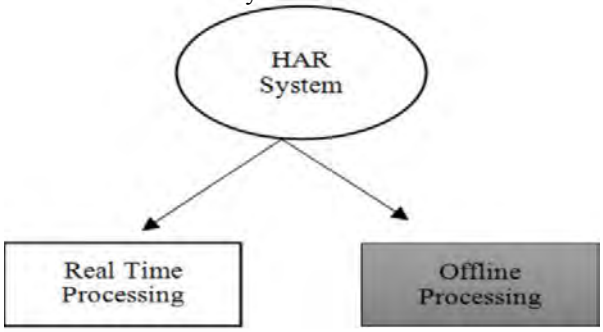
Considering the modern and hectic routine managing health without regular checkup from doctors and costly instruments seems hard but this device using HAR (Human Activity Recognition) which uses the sensor based modality to take readings for a particular amount of time and then check for alternate results for the given parameter makes it easy .

If any uncertainties are found in these readings the application will send a notification to inform the patient about their health on their watch and then it will also send a notification to the person monitoring.

This will help many people to keep a check on their family friends and provide with help as soon as possible.

II. LITERATURE SURVEY:

A Human Activity Recognition (HAR) system can automatically recognize physical activities, which is a key research issue in mobile and ubiquitous computing. An HAR system performs tasks of recognizing different human daily activities from simple to complex. The sensors involved in an HAR system can be video sensors, inertia sensors, and environment sensors. The GPS receiver can also be used for activity recognition but is limited to outdoor environments.
Bishoy Sefenet. al. [ICAART 2016] said that In order to achieve the best between the system's computational complexity and recognition accuracy, several evaluations were carried out to determine which classification algorithm and features to be used. The analysis results showed that naive Bayes performs best in our experiment in both the accuracy and efficiency of classification, while the overall classification accuracy is 87%.



Subhas Chandra Mukho padhyay [IEEE 2015] has reviewed the reported literature on wearable sensors and devices for monitoring human activities. The human activity monitoring is a vibrant area of research and a lot of commercial development are reported. It is expected that many more light-weight, high-performance wearable devices will be available for monitoring a wide range of activities. The challenges faced by the current design will also be addressed in future devices. The development of light-weight physiological sensors will lead to comfortable wearable devices to monitor different ranges of activities of inhabitants. Formal and Informal survey predicts an increase of interest and consequent usages of wearable devices in near future, the cost of the devices is also expected to fall resulting in of wide application in the society.

International Journal of Emerging Technologies in learning (iJET) 7(4):1-4 · January 2019 states Wearable sensor and internet of things technology for better medical science: A review E-health becomes one of the internet's products for healthcare. The problems of health service such as far hospital and expensive examination fees become the emergence of this technology. Consequently, people reluctant to check their health to hospital. E-health provides information on disease prevention, detecting early symptoms, and medical parameters from a far distance. Internet of things became the main concept in this system, which combines wearable sensors, communication systems, and mobile user interfaces. Reliable and valid system, easily carried, help the doctor to monitor patients from far distance expectantly to overcome the problems. The aims of this paper review are describing how an internet of things technology and wearable sensor help medical science and Find the best way to create a health monitoring system.

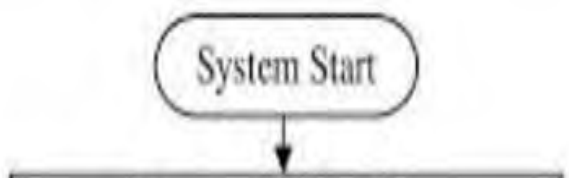
III. PROBLEM DEFINATION:

In the day to day life of a normal person it is hard to keep track of our and our loved once health. Many people need to move far away from family and in such situations taking care of someone is difficult and thus sometimes it leads to delay in providing with proper medical help. Our goal of the project is to decrease the delay and help monitor the loved once.

IV. APPROACH TO THE SOLUTION:

As many people of different age group were smart watch and like to keep a track of their day to day activities, we propose an application that count the number of steps the person is walking and also through the sensors already present in the smart watch it will also tell the pulse rate of the person this data will be available to another person who will be monitoring the person using smart watch. In case any readings are found to be fluctuated both the user and the person monitoring will be informed notified. This ensure that user is been provided with proper medical help when needed in time.

.



V. PROPOSED WORK:

In this project we try to reduce the time needed for the medical help to reach the patient. This can be achieved when someone monitors the patient or the user using the application and therefore keeps a proper watch over patient's record.
Two entities that is step tracking and pulse will be checked. The smart watch user will register upon the application and will receive a unique registration Id. Through this registration Id another person will keep a track of the user steps and pulse rate. So if the user is exhausted his pulse rate will fluctuate and hence a notification will be send on the smart watch to the user as well the person monitoring him/her. This will help the person to quickly connect with the user and check up on him, and if needed call the medical help to his /her location.

5.1 MODULE:

Module 1

Selection of the tool

We have selected a smart watch as our tool for inputs.

Module 2

Selection of simulation tool, learning the required language for the application.

Module 3

A quick notification to both user and the person monitoring.

Module 4

Designing Algorithms

Module 5

Final application

5.2 WORKING:

WORKING ON SMARTWATCH

Opening the application on smartwatch it will ask to register or monitor. On registering it will provide with a unique registration Id and will start calculating the entities. If found any fluctuation it will notify the user.

WORKING ON MOBILE PHONE

Opening the application it will ask for register or monitor. When selected monitor it will ask for the registration Id by providing the import of registration Id now the data of that registration Id can be Accessed to the person monitoring. If found any fluctuation it will notify the person monitoring.

Thus helps us keep proper check upon the user and if necessary provide proper medical help in time.

VI. SIMULATION TOOL

The core function of the project is to create an application to monitor the health of person for which the some following tools were used hardware and software.
As in hardware a low cost smart watch with basic functioning is used.

VII. HARDWARE REQUIREMENT

Model: W1, Chipset: MTK6261D,
Screen: 1.54" HD IPS Screen, Touch
Screen: Capacitive touch screen, RAM +
ROM: 64M + 32M, Resolution: 240*240

,Camera: Support, Bluetooth: Support Call: support, Size: 55.4*41*11.8mm.

VIII. SOFTWARE REQUIREMENT

The main aspect in the software part is android studio which can be used to make android application for implicit working with algorithms. Android studio work in two aspects as in two parts Front End and Back End

At Front end it contains various xml files which are the front representation of the application called as layouts. The working of those layout totally depends on the back end side work which either conation Java Files or Kotlin files. For this project we will be working with Java files.

IX. CONCLUSION

This paper introduced a low cost, low-power embedded system for HAR

Smartwatch with proper implementation made application including pedometer work and different parameters to keep a person's health in check.

This new innovative concept of the project will increase the rate of person health and keep it in check.

In this design we improved the way of tracking calories and steps in data section. This will make sure that whenever a person's health factors are differing in a unusual manner they can get help at ease through this project.

This will help in society for people who live away from their family and wants to monitor the health of their members.

X. FUTURE SCOPE:

We tend to develop this application supporting both ios and android platform.
As we tend to build this application to provide medical help in time to the people risk decreases and health issues can be handled.
This application is cheaper than to go and buy an expensive smartwatch.
In low cost we can provide proper help in time to the patient saving their lives.

Refrences:

Wearable sensor and internet of things technology for better medical science: A review: International Journal of Emerging Technologies in Learning · January 2019

Deep learning for sensor based Activity Recognition; A Survey: Elsevier 14 Dec 2017

Android Based Student Data Monitoring System

Prof. Prerana Deshmukh,Ankita Modak ,Adarsh Jais
Computer Technology
Priyadarshini College of Engineering Nagpur,India
prerana40484@gmail.com,ankitamodak15@gmail.com,yuvrajjais07@gmail.com

Shraddha Rekhe ,Ankush Kure ,Sanskruti Gupta
Computer Technology
Priyadarshini College of Engineering Nagpur,India
shraddha.15111998@gmail.com, ankush.kure@gmail.com, sanskrutis.gupta16@gmail.com

Abstract:**As of late, the Android Technology with web administrations has acquired numerous exceptional changes the versatile applicationadvancementfield.Thisapplicationgivesasummed up answer for screen the different works that are done by a College for overseeing it. The framework gives a straightforward interface to support of understudy data. It tendstobeutilizedbyinstructivefoundationsorcollegestokeep up the records of students. The creation and the board of precise, state-of-the-art data with respect to a students' scholastic profession is fundamentally significant in the college just as colleges. Well-informed Connect manages all sort of understudy subtleties, scholarly related reports, college subtleties, course subtleties, educational program, batch subtleties, arrangement subtleties and other asset related subtleties as well. It will likewise have faculty subtleties, batch execution subtleties, students' subtleties in all viewpoints, the differentscholasticnoticestothestaffandstudentsrefreshedby thecollegeorganization.Itlikewiseencouragesusinvestigateall the exercises occurring in the college. Various reports and inquiries can be created dependent on tremendous choices identified with students, batch, course, faculty, exams, semesters, and certification and in any event, for the whole college.**

Keywords— Android Technology, Administration, Maintain Records, Student Information System

I. INTRODUCTION

The structure and execution of an extensive understudy data framework and UI are to supplant the present paper records. College Staff can straightforwardly get to all parts of an understudy's scholarly advancement through a protected, online interface inserted in the college's site. The framework uses client validation, showing just data important for a person's obligation. Moreover, each sub-framework has verification permitting approved clients to make or refresh data in that subsystem. All information is altogether assessed and approved on the server before genuine record change happens. Notwithstanding a staff UI, the framework plans for understudy UI, permitting clients to get to data and submit demands on the web, in this way lessening handling time. All informationisputawaysafelyonSQLserversoversawbythe college executive and guarantees the most noteworthy conceivable degree of security. The framework includes an unpredictable logging framework to follow all clients' entrance and guarantee adjustment to information get to rules and is relied upon to expand the productivity of the college's record the executives along these lines diminishing the work hours expected to get to and convey understudy records to clients.

Fig 1: Android Technology

Already, the college depended vigorously on paper records for this activity. While paper records are a conventional method for overseeing understudyinformationanddiversework,thereareafewdown sidestothis technique. Initially, to pass on data to the students, it ought to be shown on the notification board and the understudy needs to visit the notification board to watch that data. It requires some investment to pass on the data to the understudy. Paper records are hard to oversee and follow. The physical effort required to recover, adjust, and re- document the paper records are allnon- esteem included exercises. This framework gives a straightforward interface to the upkeep of understudy data. Instructive organizations or colleges to keep up the records of studentseffectivelyverywellmayutilizeit.Accomplishingthi s goalistroublesomeutilizingamanualframeworkasthedatais dissipated,canberepetitiveandgatheringimportantdatamight beverytedious.Everyoneoftheseissuesisunraveledutilizing Smart Connect. The paper centers around showing data in a simple and clear way that gives offices, for example,

profile production of understudy, teacher, head, and more significant position authority in this manner diminishing desk work and computerizing the record age process in an instructive organization.

II. EASE OFUSE

A) Purpose

Theobjectiveoftheprojectistomakeit easyforthestudentsto accesstheirdataviatheapp.Theappwillmanageallthestudent information in the institute. It will provide the centralized control over thesystem.

B) Scope

The app is based on students and teachers point of view. The goal is to make it easy for the students to access all the data remotely. Just some clicks and the data will be at fingertips.

Thebasicapproachistouseitonallthesmartphones.Allthe students will get access to student panel and similarly for teachers andadmin.

III. LITERATURESURVEY

All researchers have aimed to develop and provide a generalized solution to monitor the various works that are carried out by a College for automation of various tasks. They provided up to date information of the system, which improved efficiencyofcollegerecordmanagementanddecreasethespace between student and college. The major contributions to this topic are summarizedbelow:

S. Shivasubramanian, S. Sivasankaran ,and S. Thiru Nirai Senthi [1] proposed one of the first computational schemes An AndroidBasedMobileApplicationtoMonitorWorksatRemote Sites.Thisapplicationprovidedageneralizedsolutiontomonitor the various works that were carried out by a construction companyatdifferentgeologicalpoints.ByusingaWebService, the data was stored in the remote database. This mobile application requires General Packet Radio Service (GPRS) or Wi-Fi technology to reach the remote database. Using data in the remote database, various reports were generated and projected as a MIS [Management Information System] web application.

Sanjay T. Attendance [2] proposed attendance Management System to generate an android application to calculate the attendance of the students in colleges and updating the result directly into the college server. The data would be stored in the smart phone if the internet connection were unavailable at that time. When the internet connection was available, then the faculty could login into their college account and update the attendance result.

Shradha S. Chawhan, Mangesh P. Girhale, and Gunjan Mankar [3] worked on MPBAS that helped lecturers to take the attendance of students using Smart-phone. Lecturers would login to the phone application, are connected to the server and take attendance using Smart-phone. After taking theattendanceinthemobile,lecturerswouldsenditovertothe server using GPRS and attendance list would be updated automatically. Lecturers would be able to edit the attendance byloginontothewebsite.Studentswouldbeabletoviewtheir own attendance as well as curriculum details. To reduce the chances of fake attendance, the project would include Location detection using GPS. In addition, email would be send to the students by the lecturers, notifying them of their regularactivities.

S. R. Bharamagoudar, Geeta, S.G. Totad [4] worked on Web Based Student Information Management System that provided a simple interface for maintenance of student information. Educational institutes or colleges to maintain the records of students easily could use it. The creation and management of accurate, up-to-date information regarding a students' academic career is critically important in the university as well as colleges. Student information system deals withallkindofstudentdetails,academicrelatedreports, college details, course details, curriculum, batch details, placement details and other resource related detailstoo.

Eiichiro Tsutsui, Kazuharu Owade, Yusuke Kondo, Michiko Nakano [5] A Proposal For A New- Dimensional Online Feedback System: Focusing On Individual Learner

Differences purpose of this study was to create a new method of assessing individual learner differences in the contexts of language learning. Their questionnaire-type items used in that system was based on SILL (Strategy Inventory of Language Learning) questionnaire items.

Namrata Shahade, Priya Kawade and Satish Thombare [6], proposed Student Information Tracking System an Android application to manage student attendance on mobile. In many colleges teachers used to take attendance manually. Main objective of this project is to add mobility and automation in the existing attendance process. This system helps teachers to takeattendancethroughmobileandkeepintouchwithstudent insomeaspect.ThisSystemallowteacherstotakeattendance, edit attendance, view student's bunks, send important documents in pdf format such as exam time table, question bank etc. and also helps teachers to inform students about the events that college was going to organize. This system also helped students in specifying bunks, deleting bunks, viewing their bunks. This system gives a prior intimation to student as soon as his attendance goes below the specified attendance deadline in the form of an alert. This system helps students to keep in touch with the events that college wasorganizing.

Yohei Kawaguchi, Tetsuo Shoji, Weijane Lin, Koh Kakusho, Michihiko Minoh proposed a system that took the attendance of students for classroom lecture. Their system took the attendance automatically using face recognition. However, it is difficult to estimate the attendance precisely using each result of face recognition independently because

the face detection rate was not sufficiently high. Here they propose a method for estimating the attendance precisely using all the results of face recognition obtained by continuous observation. Continuous observation improves the performance for the estimation of the attendance. They constructed the lecture attendance system based on face recognition, and applied the system to classroom lecture.

IV. SYSTEMARCHITECTURE

The system design consist of following modules:

a) Adminlogin:
The central level authority will be able to login in this portal. The changes can be made in student database through this portal only.
This portal has the access to the following:

i) ViewAttendance:
The attendance of every student and teacher will be visible from this portal.

ii) Add/RemoveTeacher:
The newly appointed teachers can be added and the no longer employees can beremoved.

iii) Add/RemoveStudent:
Similarly for the students, the student can be added or removed based on their admission status.

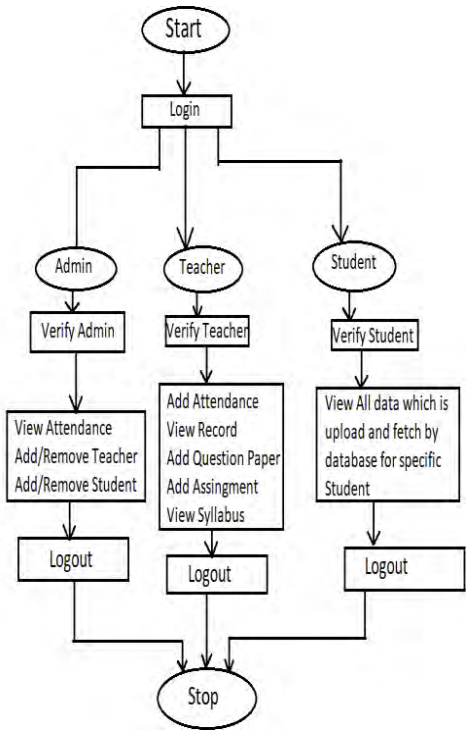


Fig: Data Flow Diagram

b)Teacherlogin:
The teachers will be able to login to this portal for taking attendance, uploading question banks, assignments, marks for the exams, and the basic student information.
i) AddAttendance:
The teacher can take attendance from the app and will be able to edit it. Total attendance will be counted.

ii) ViewRecord:
The record of the student will be visible to the teacher like student's marks and submissions.

iii) Add questionpaper:
The question paper can be uploaded through the app and will be visible to the students.

iv) AddAssignment:
The assignments can be uploaded through the app just like the question papers.

v) ViewSyllabus:
The syllabus for each and every branch will be displayed semester wise.

b) Studentlogin:
The students of the institute will be able to login to this portal and can view their personal information and attendance.Theywill beaccessibletothequestionbanks, assignments, notes uploaded by theteacher.

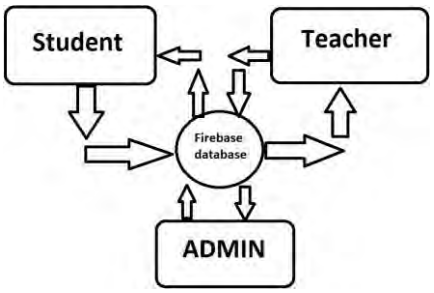


Fig: Model Flow Diagram

V. Evaluation

➢ Advantages:
- Mobility and automation will be added to the system.
- Application will be userfriendly.
- Access is granted to authorize personnonly.
- Best results are produced using multiple algorithmssimultaneously.

➢ Disadvantages:
- Designing the whole architecture can be a challengingpart.
- Tomakeitfullysecuresystemcertainencryption algorithms needs to beincluded.

VI. CONCLUSION

The framework offers unwavering quality, time investment funds and simple control. Students and their folks will likewise see results, participation and educational plan subtletiesutilizingthisapplication.Likewisestudentscansee subtleties, notices anyplace and whenever. The application will extraordinarily disentangle and accelerate the outcome arrangement and the executives procedure. It gives high security and a framework that diminishes the work and assetsrequired in customary procedure. The proposed framework gives the better approach for registering. Is playingatasks withresponsiveandappealingUI.Inthisway, based on writing study and by breaking down the current framework, we have arrived at a resolution that the propose framework will not just guide the robotization to the college, yet will likewise assist with digitizing the framework and thusly help to send assetsproductively.

VII. **Refrences**

[1] M.A. Norasiah and A. Norhayati.“Intelligent student information system”. 4th International conference on telecommunication technology proceedings, Shah Alam, Malaysia, 0-7803-7773- 7/03 2003IEEE.

[2] Jin Mei-shan1 Qiu Chang-li 2 Li Jing 3. “The Designmentofstudentinformationmanagementsystem based on B/S architecture”. 978-1- 4577-1415-3/12

2012 IEEE.

[3] Zhibing Liu, HuixiaWang,HuiZan“Design and implementation of student information management system.” 2010 International symposium on intelligence informationprocessing and trusted computing.978-0-7695- 4196-9/10 IEEE.

[4] “Mobile Phone Based Attendance System”, by Shraddha S.Chawhan1, Mangesh P. Girhale2, Gunjan Mankar3, IOSR Journal of Computer Engineering (IOSR-JCE) e- ISSN:2278-0661, p- ISSN: 2278-8727Volume 10, Issue 3 (Mar. -Apr. 2013), PP48-50

[5] DonnFelker, “Android Tablet Application Development For Dummies”, john Wiley and Sons,2012

[6] “Mobile Phone Based Attendance System”, byShraddha
S. Chawhan1, Mangesh P. Girhale2, Gunjan Mankar3, IOSR Journal of Computer Engineering (IOSR-JCE) e- ISSN: 2278-0661, p- ISSN: 2278-8727Volume 10, Issue 3 (Mar. - Apr. 2013), PP 48-50

[7] “Portable Lab: Implementation of Mobile remote laboratoryfor Androidplatform”,anIEEEpaperbyMacro Andre Guerra, Claudia Mariline Francisco, RuiNeves Madeira, Portugal.

[8] Krejcar, 0., "PDPT framework - Building information system with wireless connected mobile devices", In Icinco 2006, 3rd International Conference on Infonnatics in Control, Automation and Robotics Instinct Press, Setubal, Portugal, pp.162-167.

[9] Zigurd Mednieks, Laird Dornin, G. Blake Meike, and Masumi Nakamura, ”Programming Android”, Second Edition2009.

[10] Diego Torres Milano,"Android Application Testing Guide", Packt Publishing,2011.

[11] T. Govindasamy, “Successful implementation of e- learning pedagogical considerations,” in The Internet and Higher Education, vol. 4, 2012, pp.287–299.

[12] D. C. Hillman, D. J. Willis, &C. N. Gunawardena, “Designing online courses to promote retention,” in The American Journal of the Distance Education, vol. 8, No. 2, 1994, pp. 30–42.

[13] K. Akhila, a Novel Approach of Mobile Based Student Attendance Tracking System Using Android Application, International Journal of International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 11, November2014

[14] C. Prathyusha,” A Novel Approach of Mobile Based Student Attendance Tracking System using Android Application” Research Scholar, K.L.University, International Journal of Scientific & Engineering Advanced Research (xxxx –xxxx)

[15] Volume1 Issue12013

[16] Macro Andre Guerra, Claudia Mariline Francisco, RuiNeves Madeira, Portugal “Portable Lab: Implementation of Mobile remote laboratory for Android platform”, an IEEE paper.

[17] Mahfudzah Othman, Zainab Othman & Mohd. Norafizal A.Aziz, “Student Management System in schools: The use of web-basedapplications”,J. GADING, 10(1),pp.87-98, 2006.

[18] Newman-Ford, L.E., Fitzgibbon, K., Llyod, S. & Thomas, S.L., “A Large-Scale Investigation into the Relationship between Attendance and Attainment: A Study Using an Innovative, Electronic Attendance Monitoring System”, Studies in Higher Education, 33(6), pp. 699-717,2008.

M-Commerce Shopping Using Nfc Technology

Deepti S. Lawand, Prof. Prashant Nitnaware

Dept. of Computer Engineering

Pillai College of Engineering

Navi Mumbai, India

dlawand@mes.ac.in, pnitnaware@mes.ac.in

Abstract—**Mobile commerce (M-commerce) is the buying and selling of goods and services through wireless handheld devices such as smartphones, tablets. In traditional shopping, the customer needs to physically pick up the items to be purchased and carry cash or credit/debit cards with them to make payments M-commerce application become most popular amongst the mobile phone users who want to do their business, financial activities easily and securely anywhere, anytime. The use of physical cash is reduced tremendously after the e-cash comes in the picture. An emerging technology behind the mobile payments is Near Field Communication (NFC). Near Field Communication is a wireless technology which is a short-range, high frequency, low bandwidth radio technology that allows transferring data within a few centimeters.**

In this paper we propose an android based mobile commerce system using NFC. The system will read the NFC tag of the product and add it to the shopping cart. It would provide methods to change the quantity of products purchased and edit the cart. Invoice will be generated and NFC based payment will be done according to the generated invoice.

Keywords— M-commerce, NFC, wireless technology shopping cart

I. INTRODUCTION

In traditional shopping system the customers goes to the store and purchase the needed products and add it into the cart/trolley. The customer physically picks the product and keeps it in the trolley and carry the trolley all around till the shopping is done. Once they are done with the shopping, they need to stand in the queue for payment. Using M-commerce all this process can be done easily. M-commerce enables users to access online shopping without using a desktop computer. M-commerce include mobile banking, virtual marketplace apps like the Amazon mobile app or a digital wallet such as Apple Pay, Android Pay and Samsung Pay. In most M-commerce enabled platforms the mobile device is connected to a wireless network that can be used to conduct online product purchases [3]. In this paper, we propose an Android based mobile phone application which is integrated with NFC technology**.** The user will do the entire shopping process with the help of their Android mobile phones with a software application that would read and process NFC Tag of the products, which are to be purchased. To store products details in a shopping list/cart NFC tags of those products are are read. Users will be able to perform editing of existing products in the cart such as the process of addition, subtraction of quantity or deletion of the product all together from the cart. Users will be able to add new products, modify existing products, change the quantity or delete the product from the cart. The amount of money spent will always be known to the user and could verify the same.

II. LITERATURE SURVEY

A large population that neither had a landline nor a bank account in their names but now not only they own a mobile handset but are also well balanced to transact on their mobile. The mobile channel has provided a rare opportunity not only to leapfrog years of poor infrastructure development but also in bypassing geographical constraints to bring massive benefits and lifestyle changes to millions of under-served people across India. India is a land of many languages but only 2 percent of the Indian population prefers reading in English.[4] E-commerce stands for electronic commerce, on the internet. Customers can purchase anything right from an Insurance Policy to pen on sitting comfortably in their office or home and gift it to someone sitting miles apart just by click of a mouse Ease of internet access and navigation are the critical factors that will result in rapid adoption of E-commerce. Safe and secure payment modes are essential to popularize E-commerce in India. Though it offers many benefits to users, there are many reasons for not shopping online like are lack of trust, security concerns, uncertainty about product and service quality, delay or non-delivery of goods, and lack of touch-and-feel shopping experience [1][3][7][5][9].

Mobile Commerce (M-commerce) is the subset of electronic-commerce (e-commerce), which includes all e-commerce transactions, carried out using a mobile (hand held) device. M-commerce is the way of doing business in a state of motion [5][6][9]. M-commerce depends on the availability of mobile connectivity. We have become used to making mobile phone calls anywhere, at any time and m-commerce is built on that capability. According to IT experts, in the future, consumers will be able to shop using a handheld computing device, PDA, wearable computer, mobile phone or smart devices. In virtually any place – malls, restaurants, hotels, airports and other locations – this user will be able to receive coupons, download information, receive sales offers, and perform credit card transactions. In such future scenarios, M-commerce means that customers can shop anywhere, anytime.

III. PROBLEM STATEMENT

In traditional system the customer goes to the store and purchases the needed products and adds it into the cart/trolley. The customer physically picks the product and keeps it in the trolley and carry the trolley all around till the shopping is done. Once they are done with the shopping, they need to stand in the queue for payment. Using M-commerce all this process can be done easily and user friendly. Nowadays wireless communication technologies are in trend .One of the mostly used technology is RFID (Radio Frequency Identification), it uses the electromagnetic Field to track the information regarding the products. The products are classified by using RFID. There is an RFID tag attached to the products so by just scanning the tag in the products the details of the products is entered into the system. This system was helpful in the billing counter by making the billing service by just scanning the product to add them manually. The NFC technology comes in wireless world the NFC and RFID are related closely in the case of working but the NFC does not need any specialized device as it is inbuilt in most of the smartphones. The NFC tag is also called as smart tag or info tag as it contains the information about the particular product. They have a bit of memory storage and a radio chip.

IV. PROPOSED ALGORITHM

Step 1: Start
Step 2: Open the Application.
Step 3: Login into Application.
Step 4: If the login is successful, then scan the NFC tag of Product else go to step 3.
Step 5: After successful scan, retrieve information of product from database.
Step 6: Add the product to cart and send the list to the database.
Step 7: Place order and get the order details.
Step 8: Check if order accepted or not if accepted go to step 9.
Step 10: Receive OTP and verify it from the merchant.
Step 11: If OTP is valid, then pick up the Order else re-verify.
Step 12: End.

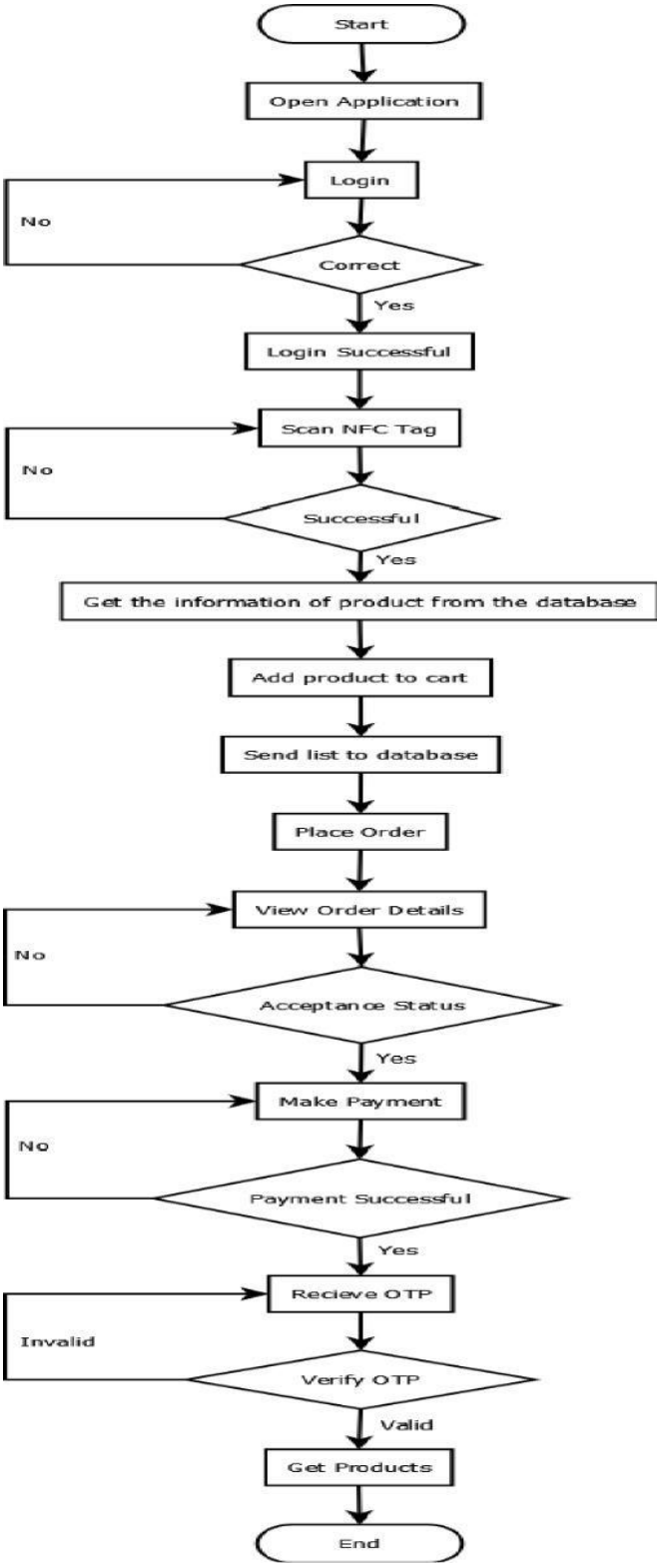


Figure 1 Flow Chart of shopping using NFC

V. RESULT AND ANALYSIS

The Proposed system provides online shopping with mobile phones with NFC technology. First it starts with sign in

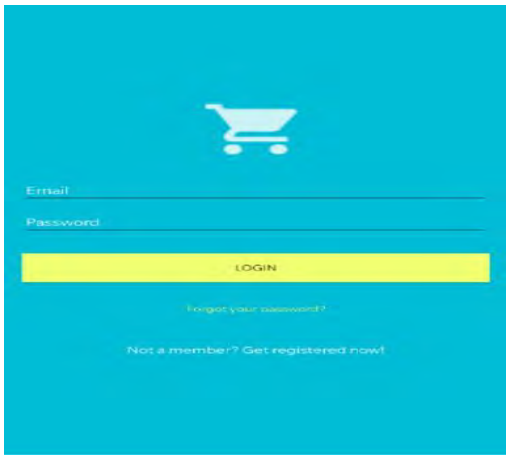


Figure 2 Login Screen

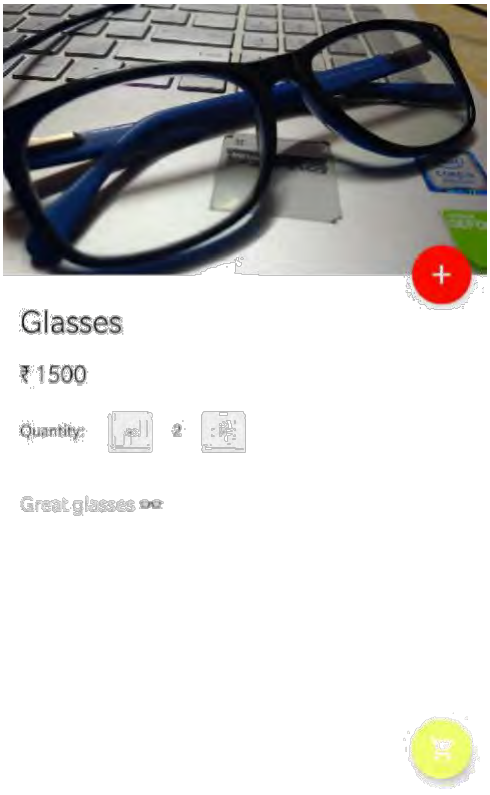


Figure 4 product Details after scanning tag

VI. CONCLUSION

The proposed system is a software prototype that provides a simplified method for shopping. It results in a fast and easy procedure saving customer time and allowing more access towards commodities. Our NFC based technique will prove to be highly efficient for shoppers to reduce the load of the goods carrying around as this application can scan NFC tags of products by NFC which is short-ranged wireless technology available in smart phones. The application could also be used for shopping by collecting users' personal information like address and let the mall or supermarket deliver the products to you home directly. In that way, the customer need not carry the goods with them, making it convenient for them. An Expert System can be integrated with application so the user's personal information regarding their health will be stored in database and then the system will suggest products related to his health.

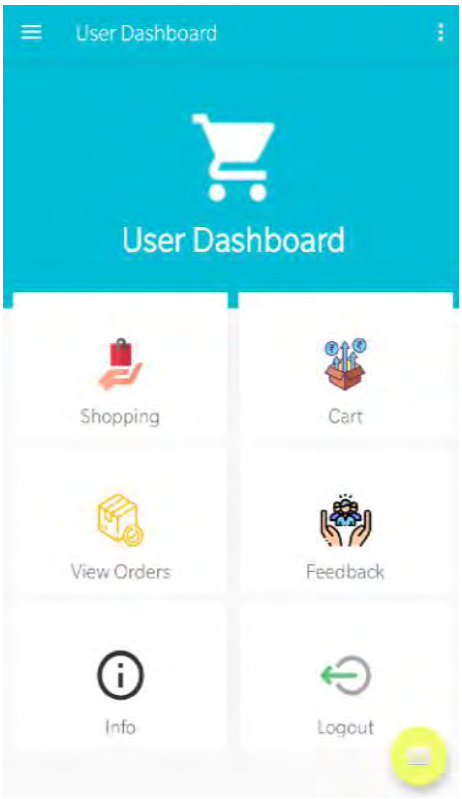


Figure 3 User Dashboard

VII. References

[1] Swati Jadhav, Sneha Kamath, Sonam Yadav, AmrutaRajput, Prof. K. S. Sakure "Smart Shopping Application using NFC" in International Research Journal of Engineering and Technology, Volume 5, Issue 3 (March 2018)

[2] Jay Dave, Sagar Gondaliya, Bhumi Patel, Arlina

Mascarenhas, Maya Varghese "NFC Based Store Automation" in International Organization of Scientific Research, Journal of Mobile Computing & Application, Volume 4, Issue I (Jan. - Feb. 2017)

[3] Margaret Rouse. "M-Commerce (Mobile Commerce)".

Internet: https://searchmobilecomputing.techtarget.com/definition/m-commerce, April 2019, [April. 5, 2019].

[4]Javrvan. "Wonderful NFC Projects to arouse your ideas".

Internet: https://www.elecrow.com/blog/wonderful-nfc-projects-to-arouse-your-ideas/, 5th September 2016, [April. 5, 2019].

[5] NFC Today, "NFC Tags Explained". Internet: https://nfc.today/advice/nfc-tags ,23rd March 2018, [April. 5, 2019].

[6] Firebase, "Services offered by Firebase". Internet: https://firebase.google.com, 12th April 2019, [April. 10, 2019].

[7] Refsnes Data, "Introduction to XML". Internet: https://www.w3schools.com/xml/xml_whatis.asp, 20th March 2014, [April. 10, 2019].

[8] Admin MindOrks, "Introduction to Android Studio IDE".

Internet: https://blog.mindorks.com/introduction-to-android-studio-ide, 17th December 2018, [April. 10,2019].

[9] Matthew Tyson, "What is JDK? Introduction to Java

Development Kit. " Internet: https://www.javaworld.com/article/3296360/what-is-the-jdk-introduction-to-the-java-development-kit.html, 9th August 2018, [April. 11,2019].

Autonomous Agricultural Pesticides Spraying Hex Copter Drone.

P. U. TEMBHARE, A. KUMARI, N. A. SHAHU, R. R. MADHUMATKE, S. KUMBHALKAR, S. B. JAGTAP

Abstract- ***India is having approx. 159.65 million hectares of agricultural land and Agriculture plays an important role in the growth of any country. The production rate of crops is getting less because of climate issues, humidity, global warming, population pressure, etc. To cover the productivity of crops we need to prevent them from the diseases. Spraying the pesticides on crops may help the crops to be safe from insects, etc. but it also affects the human's health. Therefore, to solve the problem like this agricultural pesticide spraying by hex copter drone is one of the best solutions. This paper discusses special attention to this area of agricultural, and drones with their problems and solutions and their implementation for pesticide spraying.***

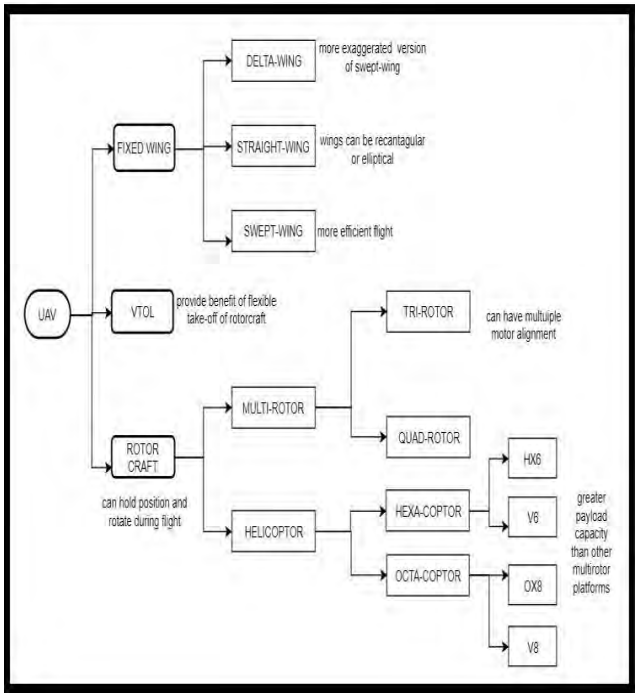
I. INTRODUCTION

In India, Agriculture is a major sector of our economy but still, it is far short of western countries when it comes to adapting the latest technologies for better farm output [1]. These technologies for spraying pesticides include Pulse width modulation improves accuracy, Spot spraying technology for herbicides, Drone spraying set to take off, Decision support tool to improve spray timing. Among these applications, the use of the drone for agriculture is called agricultural drones. Agricultural drones can be used for Soil and field analysis, Seed planting, Crop spraying and Spot spraying, Crop mapping and Surveying, Irrigation monitoring and Management.

To create better crop diversity more environment-friendly techniques are used. Due to changes in climate crops may suffer from different kinds of pests, and many farmers choose the use of the pesticide for their crops. This pesticide is used for plant protection. Farmers used to do pesticide spraying manually, resulting they harm more than just the "pests" at which they are targeted. They are poisonous, and contact with these pesticides can cause a number of health effects. They are linked to a variety of range of serious illnesses and diseases from respiratory problems to cancer. So, to avoid this a drone is used to spray the pesticide instead of human

In many developed or developing countries, they started to use these latest technologies in farming. These UAVs, if used for pesticide spraying, can reduce the workload of the farmer and also takes less time for spraying the fertilizers. It is also observed that the lower level of the plant also gets an efficient amount of pesticides because of spraying the pesticides by UAV drones. The use of agricultural drones for pesticide spraying will become easier for farmers if those drones are autonomous, i.e., once the fields length, breadth, area, or co-ordinates are set by the laptop or mobile then the drone will continue the work until any technical problem occurs like the battery is about to die, and an another solution to get back the drone if it has less battery by giving the in-build instructions suppose if it has left less battery then it will automatically calculate the distance (km) to cover and the distance(km) to get back to the place where it takes off and will come back automatically to that.

II. TAXONOMY



2.0 DRONE DESIGN OR CONFIGURATION SELECTION

After studying many designs of the drone, finally, the Hex copter configuration is selected. The Hex configuration consists of six motors and six propellers. The number of motor and propeller are the same. To reduce the manufacturing cost, a custom design is made with the hex copter frame as shown below. Hex copter configuration is selected because it has the characteristics of both, a Quad copter and an Octocopter. One of the advantages of the hex copter is the if in the mid-flight one of the components of the propulsion system fails i.e. if any one of the motor or propellers get damaged, it is capable of land safely with the 5 rotors. The Pixhawk flight controller is used for the programming of the drone because it is very much stable. The Pixhawk flight controller is also an open-source platform

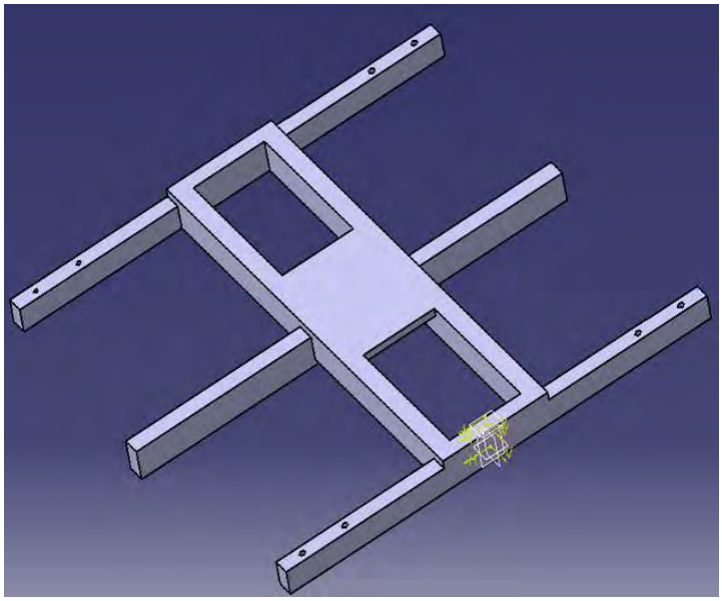


Figure 2.1.1 Hex copter frame with custom design.

2.1 OTHER DESIGN CONSIDERTIONS

Other designs we such as fixed-wing electric planes, liquid fuel UAVs, Quad copter, Octacopter were also taken into consideration for this project, but due to freeness and versatile quality of the hex copter, it is selected. Hex copter has a higher load-carrying capacity, like octocopter. Hex copter also has the accuracy of the quad copter. It can also fly in high wind also.

2.1.1 DESIGNING CONSTRAINTS

Designing constraints for the above model are specified below.

a) It should have a flight time of at least 10-15 minutes.
b) It should evenly spray the content in the field.
c) It should be working autonomously.
d) It should be carrying a sufficient amount of spraying material.

2.2 CALCULATIONS

2.2.1 BATTERY LIFE OR FLIGHT TIME

Flight time = (battery capacity in mAh) * 60

1000 * total current drawn

Ex. = 10000 * 60

1000 * 1202

= 5 min

For 1kg payload. This is theoretical calculation.
This is just a prototype for the proof of concept.

2.2.2 TOTAL FLYING WEIGHT CALCULATION

The total flying weight is also termed as the take-off weight. The total takeoff weight comprises of the weight of the all the components of the drone-like frame, battery, payload, etc. all the electronic components used example wires and connectors.

The total flying weight calculations are as follows:

Motor = 66gm * 6 = 396gm
ESC = 4gm * 6 =24gm
Propellers = 12gm * 6 = 72gm
Power distribution board = 20 gm
Flight controller kit = 150gm
Raspberry pi and camera = 160gm
Spraying mechanism = 200gm
Battery = 650gm
Airframe = 500gm

Total weight without payload = 2172gm approximately.

III. FLYING SYSTEM OR PROPULSION SYSTEM SELECTION

3.2 CUSTOMISED MOTOR MOUNTS

In this project custom made motor mounts are being used. The Motor Mounts are manufactured by the 3D printing process. The material used for the manufacturing of the motor mount is Acrylonitrile Butadiene Styrene (ABS). It is strong and light in weight as compared to others.


3.3 PROPELLERS

The Propellers we are using here is the DJI Propeller. We are required to use 6 propellers. This propeller along with the motor generates a sufficient amount of thrust required for the hex copter configuration.

3.4 ELECTRONIC SPEED CONTROLLERS OF ESC

The maximum current drawn by a motor is 24.2amp, thus we are using 35amp ESC.

3.4 POWER MODULE

A power module is used to supply the required amount of power to each component of the drone. It is also used as a stabilizer. It avoids the unnecessary shortening of the electronic component due to excess of power supply. The power module is connected with the power distribution board.

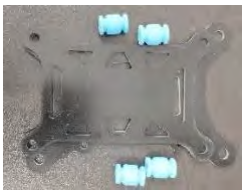
IV. CONTROLLING AND NAVIGATION MODULES

4.0.1 FLIGHT CONTROLLER (PIXHAWK)

A flight controller is basically the brain of the drone. For the proposed model we are using the Pixhawk flight controller, due to its stability and endurance. It is also an open-source and independent platform. It is also user-friendly.

4.0.2 ANTI VIBRATION KIT FOR FLIGHT CONTROLLER

It is used to secure the flight controller. It increases the stability as well. It reduces the unwanted jerks form the drones.



4.0.3 GPS MODULE

A GPS module is used to locate our drone properly. It is the main component responsible for the autonomous factor of the drone. Using GPS drones can take off and land on their own.

4.0.4 TELEMETRY

Telemetry is the connecting link between the ground station and the drone. It consists of the two components one is attached with the drone and other is attached with the ground station. A ground station can be a mobile phone or a computer. Telemetry is used to give the desired command to the drone, to set new way points, to monitor where our drone is flying.

4.0.5 LiDAR MODULE

LiDAR stands for light detection and ranging. It is basically a remote sensing technology. It is used to create a high-resolution image of the area. It is very much useful in the night for mapping and surveillance.

V. DESIGNING MATERIAL

5.0.1 ALUMINIUM SQUARE BAR

Aluminum is used for the above suggest prototype of a drone. Aluminum is selected because of its high tensile strength and ductile quality, means it bends or transform easily. It also light in weight as compared to other materials.

5.0.2 Acrylonitrile Butadiene Styrene (ABS)

Acrylonitrile Butadiene Styrene (ABS) is a thermoplastic material. It has good tensile strength and rigidity. It has excellent resistance power against stress. In this project, the motor mount is manufactured by this material.

VI. CONCLUSION

This project designed to serve a user-friendly of the technology with the farmers. This reduces the manpower required for the pesticide spraying purpose by the farmers all around. All the information and data specified above proves our concept of Autonomous spraying of pesticide through hex copter drone. This project will also have an auto take-off and auto land mode. For the take-off and landing purpose, a QR Code scanning will be done. The project is still under development phase. All the calculation specified here is theoretical. Also, we are working on to make this project to give the initial command through the Android mobile phone.

References

1. Yallappa D.1, M. Veerangouda2, Devanand Maski3 Vijayakumar Palled4, and M. Bheemanna *Yallappa D*
M. Tech. (Ag. Engg.) College of Agricultural Engineering, University of Agricultural Science, Raichur
– 584104, Karnataka, INDIA.

2. UM Rao Mogili1 and B B V L Deepak2, International Conference on Robotics and Smart Manufacturing (RoSMa2018), National Insitute of Technology, Rourkela, Orissa-769008, India*

3. Srinivasulu Reddy1, Dr. D. P. Biradar2, Prof. B.L. Desai3, Dr. V.C Patil4, Dr. Prakashgoud Patil5, Dr. V.
B. Naragund6, Varadendra Desai7, William John8, Shivprasad M Channangi9, Vishal Tulasigeri10.

the 14th International Conference on Precision Agriculture June 24 – June 27, 2018 Montreal, Quebec, Canada

*4. Dimosthenis C. Tsouros *, Stamatia Bibi and Panagiotis G. Sarigiannidis Department of Electrical and Computer Engineering, University of Western Macedonia, 50100 Kozani, Greece; sbibi@uowm.gr*
*(S.B.);psarigiannidis@uowm.gr(P.G.S.)**
Correspondence: dtsouros@uowm.com or dtsouros@uowm.gr † This paper is an extended version of our paper published in IoT4 2019 Workshop, co-located with IEEE

5. Madhukar S.C

havan, (International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-11, September 2019)

Smart Guide Stick for Blind People Using Artificial Intelligence

Divya Rathore,Damini HardePratiksha Nagrare
Department Of Computer Technology,
Priyadarshini College of Engineering
Nagpur, India

Mona, KomalMadarkar, Prof. Dr. (Mrs.) A. V. Dehankar
Department Of Computer Technology,
Priyadarshini College of Engineering
Nagpur,

Abstract:**Visual defect is a condition in which a person unable to see that causes problem, which cannot solve by medical means. The difficult situations coming in front of blind person is to distinguish obstacles appearing in front of them. Due to this, they are not able to move from one place to another. Thus, the smart blind stick helps the blind people to perform navigation and to do their work easily and comfortably.Innormalstick,thedetectionoftheobstacle is not done and normal stick is not efficient for visually impaired person. The typical blind stick is not able to identify what type of things or what type of objects come in front of him/her. The rapid advancement in technology has seen the innovation of better system for assisting the disabled, including the blind, such as AI image processing, which can provide intelligent navigation capabilities to the blind. This paper reviews the design of the smart guide stick having A.I. technologies such as object detection which helps the blind person to navigate easily.**

***Keyword: Image processing, filtering, feature extraction*.**

I. INTRODUCTION

Thevision,constructamentalrepresentationoftheworld around us. This contribute to our ability to successfully navigate through physical space and interact with important individuals and objects in ourenvironments.

Eye is the most crucial part of human body, considering that 83% of the information that a person obtain from the environment is through the sight. However, there are many people with serious visual impairment that prevent them from travelling individually.

According to a 2018 report by the World Health Organization(WHO)thereareroughly1.3billionpeopl e in the world who are visually impaired, 39 million of whom are blind, and 246 million are low vision. Accordingly, these individuals must use a range of technique and tools to help them in movement.

Some of the oldest equipment for the visually impaired havebeenthewalkingcanealsoknownasthewidestick, and the guide dog though useful these aids have significant setbacks. The rapid advancement of modern technology has, however seen the advent of better system such as the assisted vision smart glasses and smart stick that can provide intelligent navigation capability to the blind. This paper reviews the design of a smart guide stick, i.e. an Artificial Intelligence based smart guide stick, equipped with image recognition technologiesthatcapturefront,side,andbackendimage s and use AI and machine learning to process them.

Smart guide stick is mainly design to detect obstacles whichmayhelptheblindpersontotraversecarefree.The audio messages will keep the user attentive and thus reduce the chances of accidents. A voice enable automatic switching is also incorporated to help them in private space as well. The propose system detect the obstacle images which are present in outdoor and indoor with the help ofcamera.

ThesmartGuideStickisasimpleandpurelymechanical device to detect the obstacles on the ground. Thisdevice is light in weight and portable. But its range is limited due to its own size. It provides the best travel aid for the person. The blind person can move from one place to another independently without others help. The main goal of the system is to provide an efficient navigation means for the blind persons which gives an eyesight by providing the information about objects present in their surroundings.

AI based smart guide stick is an innovative stick which will help in making smart decisions about the selection

Of path that has no obstacles till a certain distance. Our blind stick involve predefined commands for a blind person by using image processing technique from camera and sensors front, left and right that will

provide theblindpersonthepredefinedmessageinordertomove over their path. The camera sense the obstacles and process the image and direct the blind person with a specified message to move in a particular direction. The message will be given in the form of voice via headphones. This helps blind person inwalking.

II. SystemDesign:

System design is the process of defining theelements of a system such as the architecture, module and components, the different interfaces of those component andthedatathatgoesthroughthatsystem.Thesystemcan be broadly classified into the input unit, the control unit and the output unit. This units are aligned on the blind stick for the perfect object and the edge detection. The stick is embedded with Sensors, Adriano Nano,Batteries, Buzzer, Bluetooth, andCamera.

If the visually impaired person is walking with his/her stick into his/her hand they will be knowing the obstacles cominginfrontofthemthroughthecamerawhichwilldo the image processing of the obstacles coming before its path.

The sensors like Ultrasonic sensor HC-SR04 are placed over the blind stick which makes it robust.

If the stick of the visually impaired person has been dropped then the buzzer will is going to buzz.

If the blind person is moving in a particular direction and in a certain path, the obstacles coming in the path of the blind person are being recognised by the camera and the audio message will be given to the blind person via Bluetooth. The stick is made artificially intelligent by image processing technique and feature extracting method.

2.1 Block Diagram:

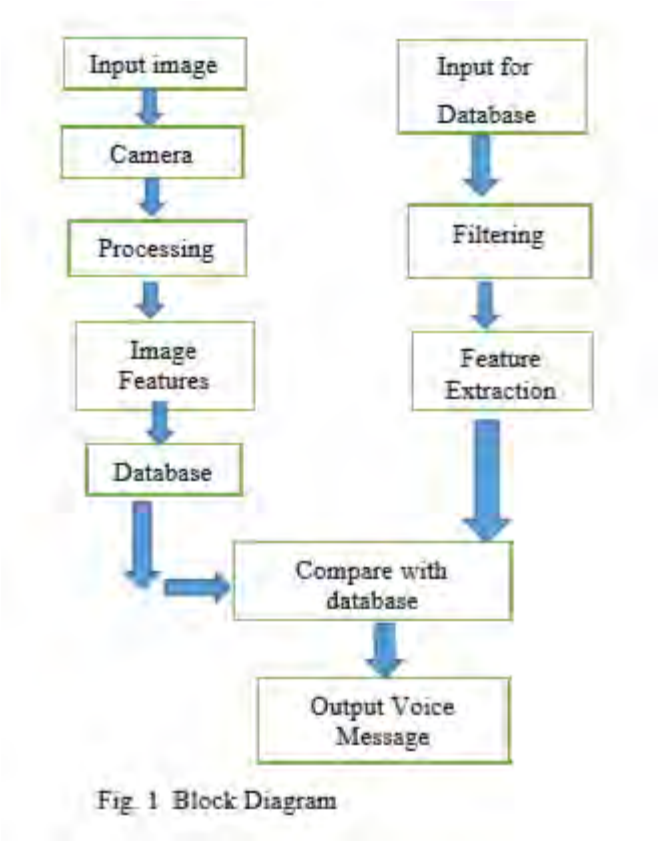


Fig. 1 Block Diagram

2.2 Camera:

A web cam is a small digital video camera directly or indirectlyconnectedtoacomputeroracomputernetwor k. Webcams are having the software that will be installed in the computer to help the blind person to record image of theobstaclescominginthepathofhim/her.Theclusterin g algorithm is used for image detection which will take the image of the obstacles as input. These input arecompared with images stored in database in order to identify them. The MATLAB is coded in C language to get the number ofimagesrecognizedbythecamerainthetextformat.Th e output is given after converting the text i.e. code in the form of voicemessage.

2.3 Ultrasonic SensorHC-SR04

The HC-SR04 ultrasonic distance sensor is a sensor used for detecting the distance to an object using sonar. The transmitters emit a high frequency ultrasonic sound, whichbounceoffanysolidobjects,andthereceiverlisten foranyreturnecho.TheHC-SR04ultrasonicsensorinthe smartguidestickwilljudgethedistanceoftheobstacleby the time taken by the ultrasonic signal to strike the obstacle and return. The ultrasonic sensor is connected with arduino. The output of these sensor is

connected by a buzzer. Arduino can be coded according to the requirements.

Fig. 2 Ultrasonic Sensor

2.4 Battery

A Battery is used for providing the electrical power to the smart guide stick. 9volt battery is used in the smart guide sticktoprovideprocessingoftheequipmentattachedover the smart guide stick.

Fig.3 : Battery.

2.5 Buzzer

Buzzer is an audio signalling device, which may be mechanical or electromechanical. The buzzer on the smart guide stick is used to indicate the blind person about the position of the bling stick when it is dropped. This will help the blind person to find his/her blind stick without any problems

2.6 Arduino Nano

The Arduino Nano is a small , complete, and breadboard- friendly board based on the AT Mega 328P i.e. Arduino Nano 3.8. The Ultrasonic Sensors are connected with the Arduino and the output is given further.

Fig. 4 : Adriano Nano

2.7 Bluetooth

Bluetooth is a wireless technology standard used for exchanging data between fixed and moving device over short distances. The Bluetooth is connected to the smart guide stick to provide information in the form of audio message of the obstacles coming in the path of blind person.

III. 3A. Image Processing

Imageprocessingtoremovenoisefromtheimageorothe r object removal, different processing unit. Here, we are using K-clustering algorithm which will help in image processing and manipulate the image to transform according to the shape andsize.

3B. Feature Extraction:

Feature Extraction is an important part in the identification of the obstacles coming in the path of theblind person. Feature extraction is used in many applications in image processing. Colour, texture, edges are the features which are used in image detection.In feature extraction here we are using K-clustering algorithm, which first extract the feature of image and after the whole implementation of process it transforms the image.

K- Means Clustering Algorithm

K-Means Clustering is a type of unsupervised learning, which is used when you have data without defined categories or groups. The goal of this algorithm is to find groups and the data, with the number of groups represented by the variable k. The algorithm works iteratively to provide each data point to one of k groups based on the features that are provided. Data points are cluster based on feature similarity. The k clustering algorithms are:

1. The centroid of the k cluster, which can be used to label newdata.
2. Each data point is labels for the each data point is assigned to a singlecluster.

Advantages:

- Obstacles detection with indicationsupport.
- Speech matching which helps blind and physical disable to recognize thedestination.
- Alert blind people about dig coming in front of them.
- Simple to use and lowcost,

IV. Conclusion:

The assistive device for blind people has been

designed focusing upon the ease of travelling for the blind people. Theprojectproposedthedesignandarchitectureofanew concept of smart A.I. based guide stick for blind people. The advantage of the smart blind stick lies in the fact that it can prove to be very low cost solution to millions of blind persons worldwide. The proposed combination of various working units makes a real-time system that makes navigation more safe and secure. The smartguidestick helps the blind people when the obstacle occurs in frontoftheperson.Thestickgivessignaltothepersonby buzzersoundwhenitdetectstheobstacle.Itcanbefurther improved to have more decision taking capabilities by employing varied types of sensors and thus could be used for different applications. It aims to solve the problem faced by the blind people in their dailylife.

References

[1] Adamo F. Schepis. "A Laser Cane for the Blind." Proceedings oftheSanDiegoBiomedicalSymposium,vol.12,2011,pp.53- 57.

[2] Borenstein, Johann & Iwan Ulrich. "The GuideCane– A Computerized Travel Aid for the Active Guidance of Blind Pedestrians."ProceedingsoftheIEEEInternationalConference on Robotics & Automation, Albuquerque, NM, 21-27 April 1997, pp. 1283-1288. [1] Benjamin, J. Malvern, NazirA.Ali, and

[3] Dhanuja, R., F. Farhana, & G. Savitha. "Smart Blind Stick Using Arduino." International Research Journal of EngineeringandTechnology,vol.5,no.3,2008,pp.2553-2555.

[4] Elmannai, Wafa & Khaled Elleithy. "Sensor-Based AssistiveDevicesforVisually-ImpairedPeople:CurrentStatus, Challenges,andFutureDirections."Sensors(Basel),vol.17,no. 3, 2017, pp. 565.

[5] Jose, Jacio, Miguel Farrajota, Joae Rodrigues, and Hans du Baf. "The Smart Vision Local Navigation Aid for the Blind and Visually Impaired Persons." International Journal if Digital Content Technology, vol. 5, 2011, pp.1-7.

[6] Kang, S. Jae, Young H. Kim, & In H.Moon. "Development of an Intelligent Guide Stick for the Blind." Proceedings of the 2001 IEEE International Conference on Robotics & Automation, Seoul, Korea, 21-26 May2001.

[7] Koley, Somnath & Ravi Mishra. "Voice Operated Outdoor Navigation System for Visually Impaired Persons." International Journal of Engineering Trends and Technology, vol. 3, no. 2, 2012, pp. 153-157. [8] Kumbhare, Shruti & A. Sakhore."ASmartStickforBlindObstacleDetection,Artificial Vision and Real Time Assistance via GPS." International Journal of Computers, vol. 7,2011.

[8] Nada,Ayat,SamiaMashelly,MahmoudA.Fakhr,and Ahmed F. Seddik. "Effective Fast Response Smart Stick for BlindPeople."SecondInternationalConferenceonAdvancesin Bio-Informatics and Environmental Engineering – ICABEE, At Italy,2015.

[9] Pissaloux, Edwige and Ramiro Velazquez (Eds.). Mobility of Visually Impaired People: Fundamentals and ICT Assistive Technologies. New York, NY: Springer,2018.

[10] Rutkowski, Leszek, Marcin Korytkowski, Rafal Scherer, Ryszard Tadeusiewicz, Lotfi A. Zadeh, and JacekM. Zurada. "Artificial Intelligence and Soft Computing."11th International Conference, ICAISC 2012, 2012 Zakopane, Poland, April 29-May 3, 2012, Proceedings, Part 1.

College Diary Social Networking Android Application

Prof. Jogi John,Prashik Meshram,Takshak Ukey
Department of Computer Technology,
Priyadarshini College of Engineering, Nagpur
Maharashtra.

Mohammed Rashid,Bhushan Kamble,Pankaj Sakhare
Department of Computer Technology,
Priyadarshini College of Engineering, Nagpur,
Maharashtra

Abstract
Given the hasty technological encroachment, social media and communication devices prolong to grow exponentially. Hence, social network has become a chief communication method in modern-day society. As a mobile application, college diary has been extensively used for personal reason as well as business dealings. Due to the lack of prose in understanding why people use college diary in emergingmarkets.Thetheoryofreasonedaction(TRA) was tailored so as to provide conjectural base to explain specific viewpoint about college diary. This provides better indulgent of mind-set and convention behavior towards college diary. Qualitative method wasadoptbymeansofpersonalinterviewinorderto bring forth belief factors about college diary.Atranscriptions were subsequently analyze using content analysis. Results indicate that behavioral beliefsaboutcollegediaryaretranquiloffivefactors, andtheyarelabeledaspersonalindulgence,features usefulness, socializing role, artifact information and entertainment.Normativeidea,inturn,ismadeupby sixfactors,andtheyarelabeledassiblings,relatives, close peers, friends in common, friends, and applicationreviewers.Thistentativestudycontributes a fundamental knowledge about college diary and social networking communication with a conjectural posture. From managerial standpoint, it also elucidatestheupshotofcommunicationcomportment andpeoplefromthesameage-grouponadoptingthe artifact.

Keywords:SocialNetworking,androidapplicationdevelopment,Marketingtool,Accesspermissionlogic,Security responsibilityfordeveloper,androidpermissionoverview.

I. INTRODUCTION:

Communicating application it becomes a useful soci networkingplatforminstantlytoindividual sandcompaniePublic media plays a major role at both personage andMoreoverMarketingisawell-improvedmethodoraction directorial levels in recent culture. With the rapid growth ofthat frequently changes the rules based on the requirement communication technologies (e.g. Internet andofpromotingandsellingproductsandUnlikethepasSmartphone), it has become a central tool in individual lifestyles and executive activities. Social networking sites, suchasFacebookandTwitter,areuniversallyknown,andhavebeenutilizeforanassortmentofpurpose.Ithasbeentraditional marketing strategy such as TVclaim that youth as juvenile as twelve have at least onecommercial and newspaper advertisement do notgivsocialnetworkingaccount.Giventheprevalentuseofsocial

networkingwithnosignofabatement,thissubjectmattermuchimpacttotheclientstheseday s.Peopleinthecurre dayareLikelytoattachtotheonlineservices. Today,socirequires continual attention and communicationprogress, we are most expected to slot in ourselves moprocess,nativeshaveslowlybutsurelyaltered themannertooelectronicdevices.Now,withtheinventionofsmatheyrelatewitheachotheranddostuff.Byutilizingsocialphones, we are in the time ofnetworkservices,activitieslikemakingacquaintancesandmobile social networking Moreover, according toEdelmainteractwithothershavedevelopintomorehandy.Thisisthe rising of these new technologies does not onlyaffebecause social network focus on initiating and buildingourlivingstylebutalsocreatedanew marketingstrategy associations, thus linking people in an almost unproblematic yet successful manner. In1 expressions of businesses, individuals and company who wanted for market expansion in the past had to goal and search for new and different markets past their intellectual zonesand familiarborders.Whilethereisnoguaranteeofformarketerstopromoteandselltheirproducts.Associ mediaisgettingmorepopular,marketershavetoquick adapttothesocialnetworkingspacetobeabletointera with customers Social media, an online community th objecttoprovideaspaceforpeopletobeintouchhasno developinto.awelllikeddirectforbrandmarketinaccomplishment, such hard works have often proved to beSubsequenttosocialnetworkingsitesuch asFaceboopricey and time consuming. College diary is a mobile application for Smartphone which is freely available in the Application Store (App Store) and Google Play. Being mainly a photo-sharing and andTwitter,Instagramcamein2010andinJanuary201 there are already 150 million assessment from researchnetworkissaidtobeavitalpartofourdays.Astechnologviewpointwiththeongoingdevelopmentof

populace by means Instagram. These social network sites do not just preseusertoexpandownnetworkbutalsopresentaplatformf thebusinesspeopletointeractwithpotentialcustome

hence, social media is moving the phenomenon of promotion.atthistime,therearemanydifferentindustries areadvertisingtheirproductandservicesfromendtoend social media. in the course of social media, big business can do well again and small business can become better known without put much venture on advertising .Social mediahasturntooneofthemostprivilegedmarketingplan usedbythebusinesscommunityaroundtheglobe.

II. SOCIAL NETWORKING

Social network is normally distinct as a web-based communication platform that aims to connect individuals and develops social relationships. It has become a pervasive interacting mechanism and the most popular onlineactivityinrecentyears.Itisyetregardasaformof tradition because users tend to use it regularly or repeatedly.Facebook,LinkedIn,MySpaceandTwitterare someofthepopularsocialnetworkingsitesthatexemplify wide-ranging,attractiveandinnovativeweb-basedservices. One of the most appealing functions about social networking is that it allows users to create public, semi-publicorprivateprofileswithinaboundedsystem,viewor share information and develop virtual connection and interactionwithothersinapragmaticmanner.Inaddition,it can help compile and organize information that is of substance

to the users, and provide links to expand contacts and connections. Over the years, it has also become an efficient channel to send or receive emails, photos, and videos, in relevant social circles, like what Yahoo and Hotmail mainly do. Given its usefulness and convenience in performing multiple tasks, it is of no surprisethatFacebookhasemergedasthemostpopular socialnetworkingsitewithoveronebillionusers.Twitteris a new admired real-time information network due to its minimalismconnectingpeoplethroughoutheworld.

College diary is a mobile-based application that enables users to take photos or pictures, apply the different manipulationtoolstotransformtheappearanceofimages, and share them instantly with friends on different social networkingsitethisapplicationtheyounger generation,and appeals to diverse societies more prevailingly than other social networking services youngsters today spend more time on Instagram than Face book and is likely because youngmobileusersareextremelydriventotakephotosor pictures using their mobile phones, and share them with others instantly. As such, the giving out of images rather than words alone has made statement with friends and broadergroupsofuserswhosharesimilarinterestsmore ideal, expedient and fascinating. On another note, by connecting physical and digital realms, college diary application enhances online presence and identity, and allows more effective interaction not only for personal reasons,butalsoforbusinesspurposes.Sinceitperforms likeasocialmediachannelwhereindividuals,companies, vendors and interest groups can easily join in, photos or pictures which are uploaded from any of them can be posted again, thus speeding up dissemination of information. As a result, a lot of established offline organizationhaveslowlybuilt-insocialnetworkingservicesinto their business in order to reach and be in touch with their actual and budding clients more efficiently.

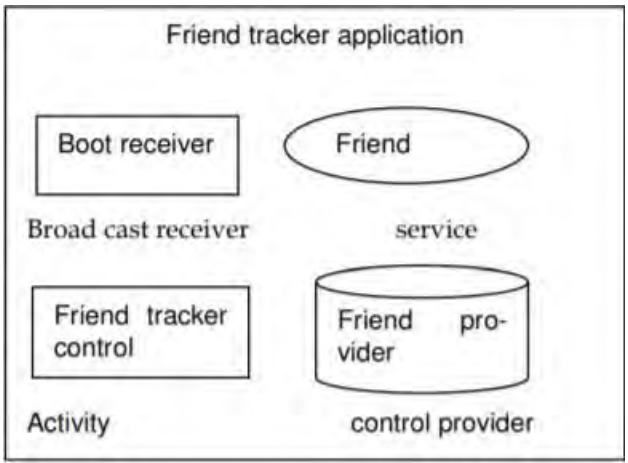
III. ANDROID APPLICATION DEVELOPMENT:

AndroidisaLinuxdisplayplaceprogrammedwithJavaand improved with its own security mechanism tuned for a mobile environment. Android join OS features like able sharedmemory,blockingmulti-tasking,Unixuseridentifiers (UIDs)andfilepermissionswiththetypesecureJavawords and its familiar class library. The ensuing security mold is muchmorelikeamulti-userserverthanthesandboxstart offontheJ2MEorBlackberryplatforms.Nothinglikeina desktopcomputeratmospherewhereauser'sapplication all run as the same UID, Android applications are individuallysoiledfromeachother.Androidapplicationsrun in split processes under diverse UIDs each with diverse permissions.Programscannaturallyneitherreadnor-write each other's data or code, and sharing data between applications must be done clearly. The Android GUI situation has some novel defense features that help support this separation. Mobile platform are mounting in magnitude, and have complex requirements as well as regulatory observance. Android supports building applicationsthatusephonefeatureswhileprotectingusers by and malevolent software. Android's process seclusion obviates the need for complex policy design files for sandboxes. This gives applications the litheness to use local code without compromise Android's security or conceding the intention additional rights. Android permissionsarerightsgiventoapplicationtoallowthemto dostuffliketakepictures,usetheGPSormakephonecall.

Wheninstall,applicationisgivenaexclusiv

eUID,andthe purposewillalwaysrunasthatUIDonthatparticulardevice. The UID of an function is used to care for its data and developersneedtobeexplicitaboutsharing datawithother applications. Applications can think about users with graphics,playmusic,andstartonotherprogramswithout special permissions. Malevolent software is an untoward realityonacknowledgedplatforms,andthroughitsfeatures Android tries to minimize the impact of malware. nevertheless,evenpoormalwarethatgetsinstalledonan Android device perhaps by pretending to be a useful application) can still temporarily wreck the user's experience).Userinthisuntowardstatewill havetoidentify and remove the hostile application. Android help user do this, and minimize the degree of cruelty possible, by requiringuserpermissionforprogramsthat dodangerous things like: – Directly dialing the calls (which may incur tolls), – Disclosing the user's private/public data, or – Destroyingaddressbooks,email,ID,etc.normallyauser's response to annoying, buggy or malicious software is simply to uninstall it. If the software is troublemaking the phone adequate that the user can't uninstall it, they can rebootthetouchtonephoneoptionallyinsafemode,which stopsnon-systemcodefromrunningandthengetridofthe software before it has a chance to run again. minimizing theconsequencesofbugs

The Android function framework services a formation on developers.Itdoesn'thaveamain()purposeor singleentrypoint for carrying out instead, developers must propose applications in terms of components. The user then uses thepalobserverapplicationtoregainthestoredgeographic coordinate and view friends on a plot. Both applications contain multiple components for the theater their respective tasks; the components themselves are classified by their module types. An Android developer chooses from predefined component types depending on the component'spurpose.



IV. SOCIAL NETWORK MARKETING

A social network site is a place for one to connect with people from their offline network and as well public they know through an online society. Look at the opportunity gained through social media to interact with latent customers, social networking sites have become a searing issue for marketers. Social network encouragement has built-in both the fixed and the non-traditional sense of marketing approach Cowden also stated that social media has not only allow the companies to interact with their patrons, but also allowing the customer to interact with other customers. It is also used to convince purchaser that the company's product or services are good .To conduct successful social network marketing, it is important to first determine who the target customers for that business are. After influential the goal customers, one can look into the most suitable platform to be in touch with the targeted customers. It is important to pay attention on what the customers want and not lay to rest them with excessive information and resources. Also mentioned that many businesses make a mistake on demanding on the quantity of their followers instead of accepting who the people that are interested in their business are. Social network marketing is a marketing involving online communities.The communities in the social networking sites are large and the gathered users often share parallel awareness. Unlike thetraditionalmarketingstrategy,socialnetw

orkmarketing requiremarketertobemoreattentivetocustomerstobuild a better image and relationship with customers Many international companies acknowledge that societal network marketing is a prospective marketing stage. According to the study, Smith school of Business, 75% of the small business has a company page on a social networking site. This has showed a momentous amount of the recognition ofsocialnetworkmarketingamongthebusinesspeople.

According to study as cited in customers these days are busierandmorepowerful,sosocialnetwork marketingcan helpthegrouptobefriendliertotheregularsatany time.In the social network promotion, the large number of social mediausershasgiventheadvantagetothecompanytoget recovered well-known in the market even though social network promotion is a potential selling strategy, there is still disadvantage for it. As the social networking siteswill allowconsumerstocommentaboutyourkindwhetheritis goodorbad;oftenthecommentofthecustomersaremore leadingandcompelling.Ontheotherhand,socialnetwork marketingdoeshaveadvantagessuchaspromotioncanbe done at almost zero cost, create huge exposure to businesses,increasingsalesandrisingthepopularityofthe brand An application developer defines one activity per "screen." actions start each other, possibly transient and recurringvalues.

V. SECURITY RESPONSIBILITIES FOR DEVELOPERS

DeveloperswritingforAndroidneedtoconsiderhowtheir code will keep users safe as well as how to deal with constrained memory, dealing out a battery power. Developersmustguardanydatausersinputintothetool withtheirapplication,andnotallowmalwaretocontactthe applicationuses,aswellasanyextensiontotheplatform anAndroiddistributionhasmade.Oneofforty **J**ournalof telecommunication,theticketsbigpicturethingto

understand about android is that every to part runs with a different UID .Typically on desktop every user has single UID and running application launch runs that program as the users UID .On Android the system gives every application fairly that every person , ties own UID For example when launching a new program the new process isn't going to run as the launcher but with its own identity its they don't know in particular if it requires a batch of permissions from a developers perception permission are just series associated with a program and its UID you can user the context class check permission method to programmatically check if a process had a particular permission like PEAD_CONTACTS this is just one of many ways permission are exposed by the runtime to developer the user view of permission is simple and consistent the idiomforenforcementmechanismManifestpermissionlike the above have a few key properties two text descriptions are required a shot t text label and longer description used on installation and icon for the permission can also be provided all permission must Also have a name which is globally unique the name is identifier used byprogrammers fo the permission and is the first parameter to context check permission permission also have a protection level thereareonlyfouprotectionlevelforpermission

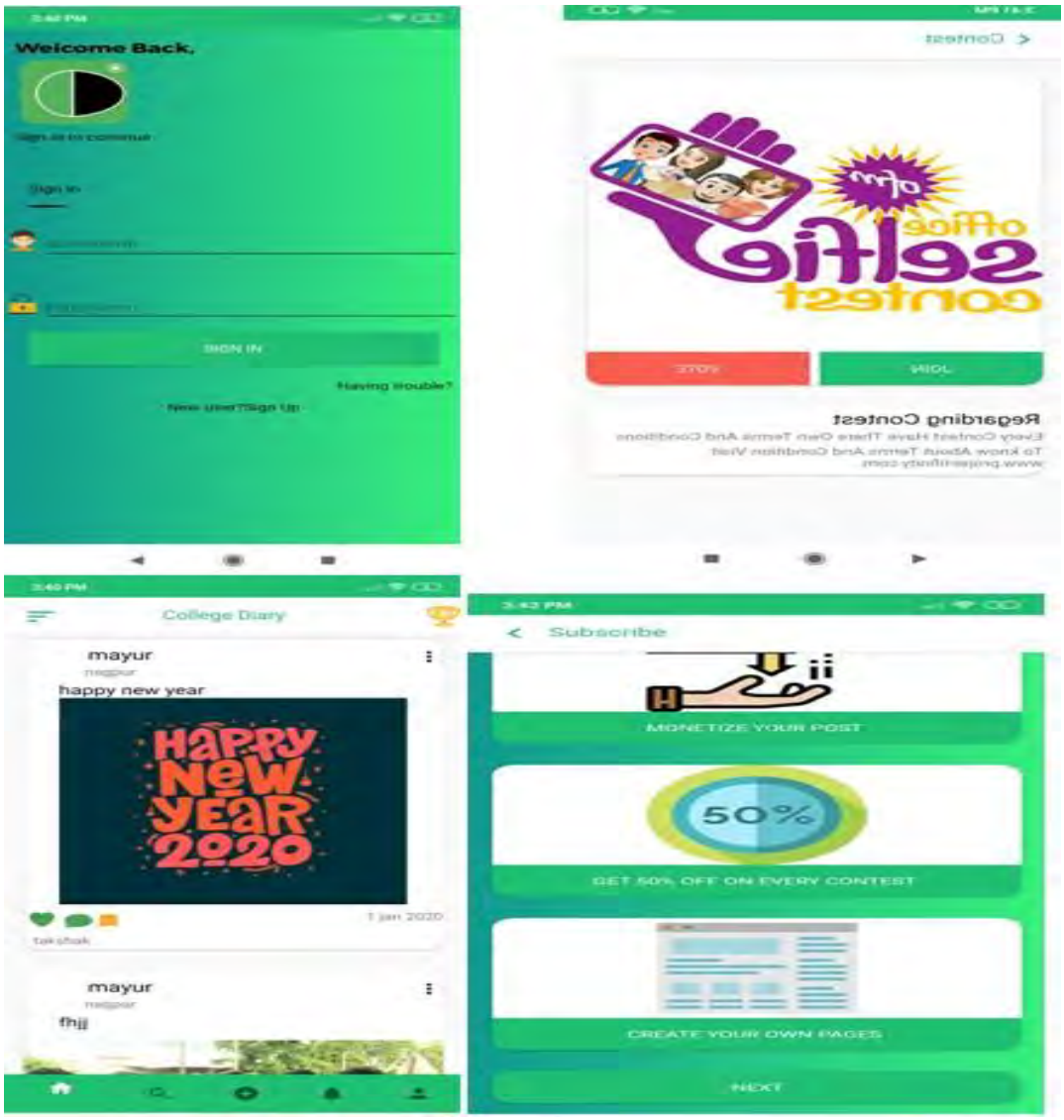
VI. ANDROID PERMISSIONS REVIEW:

Application need authorization to do thing their vendor might object to like transfer sms messages using the camera or access the owner contact database androiduser

manifest permission to follow the user agree applicationtodoanapplicationauthorizatio nneedsareexpressedinits android Manifes.xml and the user agrees to them upon install.Wheninstallnewsoftwareusershav eachanceto thinkaboutwhattheyaregoinganddecideto rustsoftware base on evaluate the developers reputation and the authorizationrequireddecidingtheupfront allowthemto focus on their goal rather than secretly while using applications permissions are sometimes called maize permissionorandroidpermissiontodisting uishthemfrom file permission to be useful permission must be associate with some goal that user understand for example an applicationneedsthereadcontentpermissio ntoreadthe users tackle book a contact administrator app needs the shouldn'tkeepingthemodelsimpleit'sposs ibletosecure the use of all the special android inter process communication mechanism with just a single kind of permission staring actives starting and receiving relay intents and invoking binder interfaces can all require the same permission therefore users don't need to recognize morethanmynewcontactmanagerneedstor eadcontact onceinstalledanapplicationspermissionca n'tbechanged by minimize the permission and application uses it minimizes the consequences of potential security flawsin theapplicationandmakesusersfeelbetterab outinstalling it when installing an function users see requested permissions in a dialog similar to the one show in install softwareisalwaysariskanduserswillinhibit edawayfrom software.

VII. RESULT:

The screenshots below show the output of the android application



Normal	Permissions for application features whose consequences are minor like VIBRATE which lets applications vibrate the device. Suitable for granting rights not generally of keen interest to users, users can review but may not be explicitly warned.
Dangerous	Permissions like WRITE_SETTINGS or SEND_SMS are dangerous as they could

	be used to reconfigure the device or incur tolls. Use this level to mark permissions users will be interested in or potentially surprised by. Android will warn users about the need for these permissions on install.
Signature	These permissions can only be granted to other applications signed with the same key as this program. This allows secure coordination without publishing a public interface.
Signature Or System	Similar to Signature except that programs on the system image also qualify for access. This allows programs on custom Android systems to also get the permission. This protection is to help integrate system builds and won't typically be needed by developers.

CONCLUSION:

Social media is a really convenient and important communicatenetworkforallthepeoplenowadays.Wecan use it to know friends and keep contact with friends that camefromdifferentcountries.

Wecanalsoshareourideassoquicklysothat allthethings could develop so fast because people could tell us their ideasandwecouldimproveitimmediately. Wecouldalso learnnewthingsonsocialmediabywatching orreadingthe thingsthatpeoplesharedontothesocialmedia.

Therearemoreadvantagesforusingsocialmedia;however, thereisalwaysadvantagesanddisadvantagesforathing. Associalmediaistooconvenientforpeople, almostmost ofthemdon'tevenhaveto'speakout'tocommunicatewith people.

REFERENCE:

[1] Babb,A.&Nelson,A.(2013,February4).Instagram and theethicsofprivacy.RetrievedFromhttp://www.scu.edu/ethics-center/ethicsblog/globaldialog.cfm?c=15267

[2] Bergstrom,T.&Backman,L.(2013).Howtheutilization of Instagram builds and maintains customer relationships.MarketingandPRinsocialmedia.Retrieved fromhttp://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-90410 Bevins, C. (2014, May).

[3] Get schooled: A visual social semiotic analysis of target'sbrandingusingInstagram.MastersTheses.Paper327. Retrieved from http://digitalcommons.liberty.edu/masters/327/Blackman,A.S.(2009,December8).Anintroduction& outline on how I will use social networking to expand mybusiness.

[4]Social Media Marketing, 6. Retrieved fromhttp://asblackman.magix.net/public/index_htm_files/social%20marketing%20paper.pdf

[5] Chianasta,F.P.&Wijaya,S.(2014,January).Theimpact ofmarketingpromotionthroughsocialmediaonpeople's buyingdecisionofLenovoinInternetera:Asurvey ofsocialmediausersinIndonesia
.

[6] International journal of scientific and research publications, volume 4, issue 1, 1. Retrieved fromhttp://www.ijsrp.org/research-paper-0114/ijsrp-p2531.pdf

[7] J.P. Anderson, Computer Security Technology Plan ning Study,tech.reportESDTR-73-

51,Mitre,Oct.1972.

Data Compression Techniques

Rishabh D Shukla , Vedant M Pandya , Shivanshu S Shrivastava
Thakur College of Engineering and Technology
Mumbai 400101 India
shuklarishabh150@gmail.com , pandyavedant18@gmail.com , shivanshu123shrivastav@gmail.com

***Abstract*— Data analysis necessarily require data compression forreductioninsizeofthedatasoastoincreasetheefficiencyof data transmission and memory management. The process of representing the information in a compressed manner so as to reduceitsmemoryrequirementiscalledDatacompression.Iti s the process of Encoding and Decoding the bits structure ofdata in such a way that it occupies less memory. Compressing data can conserve storage capacity, accelerate file transfer, and minimizecostsforhardwarestorageandnetworkcapacity.Th is paper reports the theoretical and practical nature of all the salient data compressionalgorithms..**

Keywords*— *Data Compression , Lossless Compression , Lossy Compression , Compression , Decompression

I. INTRODUCTION

Withthetremendoususageofcomputersinvariousfieldsa large number of data processing applications that require large storage memory and high speed for data transmission are increasing at a huge rate. For instance , a scheme for compressing a video may require an expensive hardware for decompressing the contents of a video[1].Thusdevelopers felt a need for various compression algorithms that will help in minimizingthesizeofdatainordertousememoryefficiently and to ease the process of data transmission. Reducing the size of a file to its half may double the amount of data to be transmitted in that space. This lead to introduction ofvarious schemes that will help in data compression.

II. BODY

Data compression is an operation of taking raw information as input , processing it through various componentsthat involve compression and decompression algorithms and producing a compressed output of the entered data. The entered data can be of the form : text, audio, picture, video, etc. Storage and transmission of a bulk of redundant data increases the space and time complexity of the system. Itcan be interpreted as a process that reduces the complexity ofthe data and thereby produces a non redundant output. In Computer Networks , Data Compression is performed in the 6th layer i.e. the Presentation layer of the 7 layer OSI Model. Along with the input and output blocks , The process of data compression depends on the following twocomponents:

Encoder:

Encoder block consists of an algorithm that is used for accepting the raw data as input and converting it into a compressedonebyreducingthesizeofthedata.Basedonits functions , it is also knoxwn as aCompressor.

Decoder:

Decoder block consists of an algorithm that is used for reconstructing the original message and providing the output without any noticeable loss of the data. Due to its functions , it is known as a Decompressor.

MEASUREMENT PARAMETERS:

Performance of a compression algorithm majorly dependson the measurementparameters.Various measurement parameters are asfollows:

Compression Ratio:

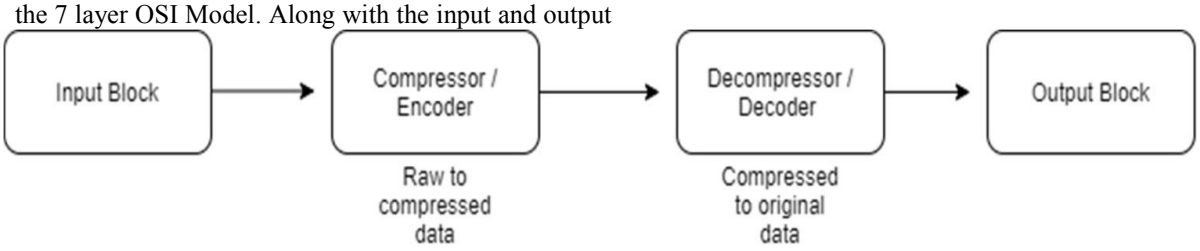
Compression ratio is basically the quantity of compression done byanalgorithm. Compressed ratio = Size of compressed file / Size of originalfile
Compression Factor : Compression factor is the inverse of compression ratio[4]

Compressed factor = 1 / Compression ratio
Compressedfactor=Sizeoforiginalfile/Sizeofcompresse d file
Compression time:

The amount of time taken by an algorithm to compress the given data is called as compression time. Unit of compression time is milliseconds (ms).

Decompression time:

The amount of time taken by an algorithm to decompress or retrieve the compressed data completely is called as decompression time[4]. decompression time is milliseconds (ms).
Saving Percentage:



The amount by which the size of a file is reduced after compression is called Saving percentage [4] Compression algorithms are used to eliminate the redundant data. Over a period of time , various data compression algorithms have been developed. There are basically two types of algorithms

:

Lossy algorithm:

Lossless

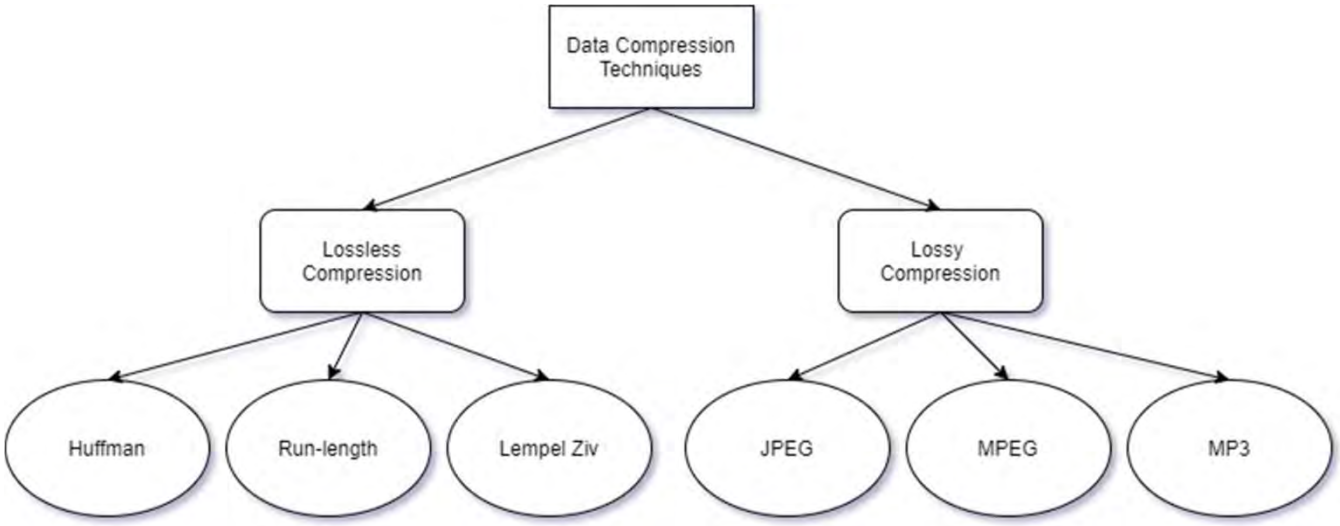
algorithm doesn't directly eliminates the redundant data, rather it temporary terminates the unuseful data.[3] The original data can be recovered or reconstructed exactly. Quality of data is maintained throughout the compression. This algorithm is mainly

Lossy algorithm directly eliminates the redundant data.

The original data can not be recovered or reconstructed exactly.[2]Quality of data degrades but it is undetectable by the user.This algorithm is mainly used for compression of images , audio or video i.e. jpeg, mp3, mpeg.

Losslessalgorithm:

used for compression of text files , programs or softwares. Data compression in lossless algorithm is comparatively less as compared to Lossy algorithm. These two algorithms can be further classified as follows:



Huffman Coding Technique:

Huffman coding technique is a type of lossless data compression algorithm where the exact data can be retrieved after decompression. It is based on binary tree frequency sorting method. In this technique, we first arrange the input data in a descending order of their frequency count. Then, according to algorithm, the least two frequency counts are added and a new node with the new frequency is generated. Wecontinuethisprocessuntilonlylastnodeispresentinthe list. The characters with comparatively higher occurance (frequency) are assigned with lower number of bits and the ones with lower frequency are assigned higher number of bits[5].

Algorithm:

S1: Create sorted nodes in descending order of their frequency count.

S2: Loop - Find two nodes (as node 1 and node 2) with the smallest frequency count.

S3: Create new node , weight of new node = weight (node 1)

+ weight (node 2)

S4: Insert new node back to the sorted list

S5: Repeat the loop until only one node is present in the list.

Advantages of Huffman Coding Algorithm:

The Compression ratio of Huffman Coding algorithm is less as compared to other algorithms[6] Saving percentage is less in Huffmanalgorithm

Disadvantages of Huffman Coding Algorithm:

In case of a fixed length code, the algorithm is inefficient as it uses more number of bits. In case of a variable length codeit becomes harder to encode and

decode as number to bits assigned to each bit may vary.

Run - Length Coding technique:

Run - length coding technique is the simplest data compression technique. This technique is basically used
when a repeated data set is entered by an user. It combines the repeated data as data value along with a counter. The counteristhenumberoftimesthatdatavalueisrepeated.Due to this , the number of bits of a large repeated data set is reduced and data compression is doneefficiently.

Algorithm:

S1: Check the current data value with the next value.
S2: If the value is same , return data value along with a counter that indicates the number of repetition.[7]
S3: If the neighbouring value is not identical , proceed to the further data value of the data set.
S4: Repeat the above processes until all the data elements of the set are being traversed.

Advantages of Run- Length CodingAlgorithm:

Itisthesimplestalgorithmandthusimplementationiseasier . For repeated data items in a data set, Run - length algorithm is quiteefficient.

Disadvantages of Run- LengthAlgorithm:

The efficiency of this algorithm depends on the number of repeateddataentriesintheinputdataset.Incaseofadataset in which the data values aren't repeated, this algorithm fails. Instead the number of bits after compression increaseswhich leads to a complete failure of the algorithm.

Lempel - Ziv Coding Algorithm:

In Lempel - Ziv Coding algorithm , a dictionary is assigned to a single character string that corresponds to all possible input character[8]. This algorithm works on continuous scanning of sub-strings that are not in the dictionary. The entered input is divided into unique sub-strings that haven't occurred before and corresponding location is assigned to each string. Then sub-strings containing previoussub-strings as their subsets are written as their location along with the remaining part of the string. In the end everything is expressedinbinaryanddatacompressionisdoneeffectivel y. Algorithm:
S1:Theenteredstringisseparatedintonumberofunitstring s which haven't occurred before.
S2: After each separation , consider input string and find the shortest string which isn't marked before.
S3: Each separation is given a location
S4: This encoding is done by assigning the location of data item and value of last bit
S5: Further , encoding is done by representing the numerical locationinbinaryandthedataitembyitspreviouslyassigne d location

- Advantages of Lempel - ZivAlgorithm:

This algorithm can complete the compression in one pass. A dynamic codeword table is built for each file.

- Disadvantages of Lempel - ZivAlgorithm:

Creates entries in dictionary that may never be used. Useful for a large amount of text where data redundancy is quite high.

III. DRAWBACKS

Every system always has somedrawbacks

- The vascular structure can be influenced by physical activities , temperature , ageing as well as by diseases buttill now there is no empirical evidence for thesame[11].
- The free flow access is denied because it can either be touch based or can be accessed from a distance and not both at the same time in current available commercialsensor.[12]

IV. RESULTS

In this paper we came across various data compression parameters. These parameters are used to define the efficiency of various algorithms. We studied about the two majordatacompressiontechniquesthatisLossyandLossl ess algorithms. These algorithms were compared with respect to eachother.Theyservevariouscompressionpurposesanda re usedeffectively.Further,Varioustypesoflosslessalgorith ms were overlooked and a comparative result was generated as the advantages and disadvantages of the algorithms with respect to each other.

V. CONLCUSION

Data compression technique is topic of prime importance in today's generation due to an increasing advancements in technology. data compression is studied from past 4 decades[4] . this paper gives an overview of various data compressiontechniques. these algorithms are then being compared with each other and results are generated.

VI. FUTURESCOPE

Veinscannersystemcanhavemoreresearchinordertohav e contact and contactless security access simultaneously . Algorithmscanbedevelopedtodetectthecauseofchangei n the veins pattern based on the change as it would help the system to know whether the change is natural or the input is manipulated.

ACKNOWLEDGMENT

We acknowledge thanks to TCET for providing such a platformwherestudentscanexpresstheirideasrelatedtoany particular topic and develop theirskillset.causeofchangeintheveinspatternbasedonthechangeasit would help the system to know whether the change isnatural or the input ismanipulated.

VII. References

1. Sethi, Gaurav, et al. "Data compression techniques." International Journal of Computer Science and Information Technologies 5.4 (2014):5584-6.

2. Mathur, Gunjan, Rohit Mathur, and Mridul Kumar Mathur. "A comparative study of various lossy image compression techniques." ETRASCT 14:165-169.

3. Mathpal, Dipti, Mittal Darji, and Sarangi Mehta. "A Research Paper on Lossless Data Compression Techniques." International Journal for Innovative Research in Science & Technology (IJIRST) 4.1 (2017):190-194.

4. Parsa, Ali. "International Journal of Application or Innovation in Engineering & Management (IJAIEM) Web Site: www. ijaiem. org Email: editor@ ijaiem. org." (2018)Types of data compression | Media Signal Processing Presentation 2| Jahanzeb Farooq , MichaelOsadebey

5. Fotie, Jean, et al. "Lupeol Long-Chain Fatty Acid Esters with Antimalarial Activity from Holarrhena f loribunda." Journal of Natural Products 69.1 (2006):62-67

6. Ramya, K. A., and M. Pushpa. "Comparative Study on Different Lossless Data Compression Methods." IJSEAS 2.Issue-1(2016).

7. Hardi, S. M., et al. "Comparative Analysis Run-Length Encoding Algorithm and Fibonacci Code Algorithm on Image Compression." Journal of Physics: Conference Series. Vol. 1235. No. 1. IOP Publishing,2019.

8. Dheemanth, H. N. "LZW data compression." American Journal of Engineering Research (AJER) 3.2 (2014):22-26

A Survey on Query Optimization Techniques in Different Environment

Riddhi Doshi,Vivek Kute
Department of Computer Engineering
St. Vincent Pallotti College of
Engineering and Technology
Nagpur, India
riddhidoshi1996@gmail.com,vivekkute@rediffmail.com

Abstract—**Database is a diversified field which has increased extensively. This extensive increase has elevated the complexity of queries. The processing time of query increases with the complexity of the query. This processing time is minimized by optimizing query. Cloud computing possesses facilities like repository, database, networking and significantly more. With focus on enhancing the execution in cloud demands the optimization of data transit time and optimization of resource rent time in cloud environment. In traditional database, query optimization is intended with one - dimensional objective whereas in distributed database query optimization is intended with multi – dimensional objectives. There are numerous optimization strategies such as randomized, static and dynamic; however these strategies require a prior knowledge of entire system. In this paper we have briefly described various methods for query optimization and proposed query optimization models for traditional and distributed databases.**

Keywords—Query Optimization, Relational Database, Distributed Database, Cloud Database, Cloud Computing, Re-Optimization, Neural Network, Query Processing, Materialization, Sub query Elimination

I. INTRODUCTION

Query optimization is the method of producing an optimal execution plan for a specific query where optimality is contingents confunction to be diminished. The fundamental aim of query optimization is to determine an execution approach that is practically optimal [1].Query is optimized in two stages:

STAGE 1: First stage is to produce the search space.

STAGE 2: In second stage an ideal plan from the Search space is identified [2].

There are three fundamental components of query optimization: access method, join criteria and transmission cost. The query optimization process is performed in three parts. First part is formation of search space; this is the accumulation of alternate execution to represent the input query. Second part is picking up the search strategy, search space is explored and the ideal execution plan is determined. Search strategies are comprehended as exhaustive strategies and heuristic strategies. Exhaustive strategies develop all feasible plans by the creation routine. BFS and DFS are frequently applied as search strategy. Heuristic strategies are noted as domain independent that confines the creation of query evaluation plan in every single point to a selection. Third part is creation of cost model, this predefines the cost of given plan which incorporates the elements such as transmission cost, memory storage cost, computation cost and so on [1]. There are two fundamental approaches for query optimization:

1. First approach is to make use a rule based or heuristic based approach for sequencing the actions in query execution strategy.
2. Second approach methodically considers the metric of various execution strategies and selects the least cost solution [1].

Different techniques used for generating adequate query execution plans are:

- Heuristic optimization
- Syntactical optimization
- Cost based optimization
- Semantic optimization.

Heuristic optimization is a rule based approach for generating better execution plans for query. Syntactical optimization is based on user intelligence of both the primary database schema and the division of data stored among the tables. The optimizer aims to enhance the capability of these joins by recognizing indexes that are effective for data retrieval. Cost based optimization requires particular information regarding stored data. The main purpose is not to generate the ideal execution plan for fetching the specified data, but –is to create a logical execution plan. In semantic optimization, optimizer utilizes the understanding of system controls to clarify or to overlook particular query if it is assured to yield an empty set. Query execution operator in particular selection operator, join operator, set operator, aggregate operator and many more assist to optimize specified query. This paper reviews different methods for query optimization and proposed models. Finally we summarize the paper along with future work.

II. QUERY OPTIMIZATION METHODS IN TRADITIONAL DATABASE

A. Generalized method for Query Optimization

Author in paper [3] briefs about generalized method for query optimization. In generalized method,

optimizer identifies the most ideal approach to execute a SQL statement after examining many elements associated to the object mentioned and situation prescribed in query. This generalized illustration comprised of query transformer, estimators, plan generator and dictionary. In this method parsed query is provided as input to the query transformer. Parsed query is described as a set of query blocks which can be embedded or connected. The function of query transformer is to identify whether it is beneficial to modify the form of query so that it allows creation of best query plan. The function of estimator is to evaluate the overall cost of given plan. Optimizer produces three types of metric named selectivity, cardinality and cost. Selectivity portrays a fragment of rows from row set. Cardinality portrays the number of rows from row set. Cost represents the work or resource used. The function plan generator is to examine different feasible plans for a particular query and selects one with lowest cost. [1]

B. Traditional Methods / Query Optimization in Relational Database

Traditional methods for query optimization are broadly used in relational databases. There are five fundamental methods as follows:

- Elimination of redundant evaluation
- Continuous or Iterative queries or results
- Caching intermediate queries or result
- Materialization
- Pipelining

Some of these methods work on reducing communication cost, some focus on reducing execution time for a given query and some focus on utilization of system resources in cloud [2].

I) Elimination of Redundant Evaluation:

Composite queries are made up of common sub – queries or common sub – expressions. Assessing these types of frequent queries only once will reduce evaluation time. Systems named IGNITE was developed, it removes redundant evaluation. This system operates on start – fetch wrapper with request window scheme. System transmits the common sub-queries to the identical data source so unnecessary answers within the queries are abolished. Source wrappers in which multiple queries comprising of common sub-queries are wrapped as a whole are decoupled by IGNITE system and execution engine provides the sub-queries to the identical source which improves data sharing within sub-queries. Authors worked on similar system to reduce the communication traffic by rebuilding the initial sub-queries to alternative sub-queries. A model was drafted where parallel query processing system accomplishes multiple queries scheduling to minimize the response time. The queries are fragmented into analogues subtasks and execution plan is reorganized maximizing reutilization of cache data [2] [3].

II) Continuous or Iterative Processing:

This method is completely distinct from other optimization methods. Principle purpose of this method is on the accumulation and propagation of data statistics prior to query execution. An approach was proposed that constantly check query execution, gather the run-time score and modulates execution plans as the query executes. Whenever novel run-time statistics are obtained, query optimizer is activated. It is complex to determine better quality statistics on intermediate computation results. Another method was proposed, which overcomes the above issue. This method produces several plans at compile time. One plan from these is adopted to execute when unfamiliar quantities are detected at the run-time with the help decision tree mechanism. It is complex to determine how many plans should be produced and retained at compile time. Progressing query optimization (POP) is another method that identifies the cardinality evaluation defects in mid execution. It contrasts the evaluated cardinality values with the actual run-time score. If inconsistency occurs, the current plan is re-optimized [2] [3].

III) Caching Intermediate Queries or Results:

Query can be optimized by caching intermediate results in single sliding window. Author proposed a model that optimizes overlapping queries with common sub-queries with the help multiple sliding windows. Implementing and temporarily storing that is caching the common sub-queries at the site with minimum communication and transmission cost and transporting the result to the site that require them as input are considerably crucial element that enhances overall execution time. Some of the proposed systems based on caching intermediate queries are EARL, Blink DB, Chive, NOVA, CBP and REX [2] [3].

IV) Materialization:

Query optimization can be obtained by considering materialized view for aggregated or clustered queries across the vast relation. Positive outcome or success of this approach relies on powerful view matching algorithm. A system was proposed based on materialization where physical functions are requited with entities present in cache during the course of run – time which removes the necessity for changes to the query optimizer. ReStore is another proposed system that makes the use of result of Map Reduce jobs which are mentioned in analytical query languages such as pig.

The output effects of Map Reduce jobs are retained to acknowledge future reutilization probability by following future jobs. An algorithm was suggested to identify either a component or all of a query can be estimated from materialized views. [2]

V) Pipelining:

By using Map Reduce the entire data processing expense for joins elevates the intensive level of work. In pipelining approach, task consisting functions is distributed into several tasks. One task reads the results of pervious task for continue processing. Nested Triple Group Data (NTGD) is a proposed model that reduces overall execution cost by minimizing the numbers of map reduce cycle. Another proposed model named AQUA (Automatic Query Analyzer) was developed with an aim

to minimize the generation of excessive intermediate results so as to diminish network and storage costs. This model makes use of two phase optimizer:

- PHASE 1: This phase creates groups of join operators for minimizing total number of jobs in map reduce so that query can be evaluated.
- PHASE 2: This phase binds intermediate results of group to produce final query results. [2]

III. QUERY OPTIMIZATION METHODS IN CLOUD DATABASE

A. Query Optimization Methods for Cloud

Cloud computing provides paid utilities for customers like repository, servers, networking, databases and software on demand. Cloud computing allows customers to hire a large number of resources for a short period of time so that complicated queries can be executed efficiently on huge amount of data. [3].In distributed environment such as cloud, data is distributed to several number of site. This data is stored in its integrality on every site or distributed on various sites. In cloud query is handled and optimized in distinct manner [4]. In distributed database, effectiveness of query is obtained by incorporating a pure key value data model [4].

In distributed database query processing involves data transmission from one site to another site across a network. The distributed database enables the association of data from various distributed sites with the help of queries [3]. These optimization methods in traditional databases are not applicable for cloud distributed database systems because they provide users on-demand services as and when needed by the users. Cloud database systems are reliable for ensuring the users with a viable query optimization solution to give the query outputs that fulfill the user constraints as well as reduce the various costs included in query execution. During query optimization, time and monetary costs are determined depending on data statistics that is readily available at that instance with query optimizer. These statistical data may be indefinite and unreliable. Therefore the query execution plan produced before the query is processed may not be the optimum one. So, adequately optimizing the query execution plan (QEP) while executing query to utilize more precise statistics will generate optimum QEP selection. Prevailing techniques either concentrate on optimizing queries depending solely on time, which is inadequate for cloud database, or do not include query re-optimization for more optimum statistics. [5]

Different approaches are described below.

I) APPROACH – 1:

This is an adaptive approach for optimizing queries based on query execution time and monetary costs. The approach is explained as follows:

1. Develop several query execution plans with distinct time and monetary costs.
2. Now based on time and monetary cost estimation each query execution plan is allotted a score with the help of Normalized Weighted Sum Algorithm (NWSA).
3. Select the query execution plan that satisfies both user limitations of time and monetary cost by applying NWSA model.
4. The evaluated execution time and monetary cost of every execution plan are systematized with the help of user-specified maximum value for every target.
5. The score is the aggregation of this systematized value
6. The query execution plan with the least score is identified and executed in phases.
7. After a phase is executed, the data statistics are updates and utilized to re – optimize the other phases until all phases are executed and the query is executed completely [5].

II) APPROACH – 2:

1. Here the execution of query is terminated for numerous times at instance defined by re-optimization strategy.
2. At each instance, a new evaluation of the cost to execute the rest of query is done with the data statistics taken from the finished query parts of the query.
3. The parts of the query which are not executed are amended with the new estimation.
4. The amended query employs more optimum estimations [5].

III) APPROACH – 3:

1. This approach proposed an adaptive optimization algorithm which actively adjusts the number of nodes.
2. In this approach initially, optimizer estimates the terminating time for recent query frequently.
3. Whenever the optimizer identifies that terminating time violated the service level agreement (SLA) between user and cloud service provider.
4. When the violation have been identified the query will be re – optimized and a new computational node will be included to contribute in query execution.
5. If there is no violation, query goes on to be executed [5].

IV) APPROACH – 4:

In this approach the re-optimization of queries is done during run-time. The approach is as follows:

1. The part of query which is executed will be diminished from the search space and a new query execution plan will be developed depending on unfinished part of the previous query.
2. This new query execution plan will be included to the search space and algorithm will find the best optimum query execution plan from new search space [5].

B. Query optimization using Machine Learning (Neural Network)

Components of Query optimizer involves the large numbers of hand – tuning. To overpower this

shortcoming machine learning can be applied to data management. Neo is the proposed system based on neural network. Neo means neural optimizer. It is an innovative algorithm based on learning that depends on deep neural networks to develop query execution plan.

Neo bootstraps its query optimization model from currently available optimizers and keeps on learning from incoming queries, building upon from its successes and learning from its failures. This model certainly adapts the fundamental data pattern and is persistent to estimation errors. Based on an experiment, when neo system is bootstrapped from a simple optimizer like postgreSQL, it makes use of traditional query optimizer as a source of proficient demonstration. Neo uses this to bootstrap its basic strategy. This technique is called as "learning from demonstration" considerably accelerates the learning process by minimizing training period from days to only few hours. Neo model is comprised of following elements: [6]

- Expertize collection
- Model building
- Featurizer
- Plan search
- Value model

Expertize collection: This is the first phase of the neo model. This phase primarily creates the experience from the traditional query optimization. This phase is only used to develop query execution plan (QEP) for every query sample workload. These query execution plan following their waiting times are added to Neo's experience. This experience is used as the base during the training phase [6].

Model Building: This is the second phase. In this phase Neo produces an initial value model from the experience gathered by the above expertize phase.

This value model is based on deep neural network which has an architecture planned to estimate the final execution time of a given partial or complete execution plan for a given query. This model is trained using supervised learning technique. This phase includes modifying each query submitted by user into features that are beneficial for machine learning model [6].

Featurizer: This model modifies each query into features beneficial for machine learning model. These features contain both query – level information and plan – level information. Neo model works well with several featurization techniques. Neo's value model uses tree convolution to process tree structures query execution plan [6].

Plan search: Once the query – level information is computed, Neo makes use of value model to search over the space of query execution plans and determines the plan with the minimum estimated execution time. Neo model executes best – first search in search space, with the help of value model as heuristic. A complete plan created by Neo, which includes join ordering, join operators and access paths is sent to the underlying query execution engine [6].

IV. Future Work and Summary

This paper contains the detailed description about different approaches and methods for query optimization. The main objective of query optimization is to minimize the transmission cost. Query can be optimized in two stages by creating search space and selecting optimal plan from the search space. The paper is divided into two main sections. First section briefs about query optimization and second section specifies different methods for optimization in different environment. In relational or centralized database query can be optimized by eliminating resultant evaluation, iterative query processing, caching intermediate results, materialization and pipelining. These techniques are designed to target traditional database system with one – dimensional optimization objective. Cloud database systems are responsible for providing the users with a feasible query optimization solution to deliver the query results that satisfy the user constraints as well as minimize the multiple costs of query execution. Query optimization in cloud database can be done by optimizing pre-optimized query called as re-optimization. Re-optimization is combined with different existing methodologies producing the new technique

References

[1] Archana Bachhav, Vilas Kharat, Madhukar Shelar, "Query Optimization for Databases in Cloud Environment : A Survey," International Journal of Database Theory and Application, Vol.10 No. 6, pp. 1-12, 2017.

[2] Tejy Johnson, Dr. S. K. Srivatsa, "A Study on Optimization Techniques and Query Execution Operators that Enhances Query Performances," International Journal of Advanced Research in Computer Science, Vol. 3 Issue 3, June 2012.

[3] Archana Bachhav, Vilas Kharat, Madhukar Shelar, "Novel Architecture of an Intelligent Query Optimizer for Distributed Database in Cloud Environment," Journal of Advanced Database Management and Systems, Vol. 5 Issue 2, September 2018.

[4] Ruchi Nanda, "Review of Query Processing Techniques of Cloud Databases," Suresh Gyan Vihar University Journal of Engineering and Technology (SGVUIJ Engg. and Tech.), Vol. 1 Issue 2, 2015.

[5] Chenxiao Wang, Zach Arani, Le Gruenwald, Laurent d'Orazio, "Adaptive Time, Monetary Cost Aware Optimization on Cloud Database Systems," International Conference on Big Data, IEEE 2018.

[6] Ryan Marcus, Parimarjan Negi, Hongzi Mao, Chi Zhang, "Neo : A Learned Query Optimizer," IEEE 2019.

[7] A. Regita Thangman, DR. S. John Peter, "An Extensive Survey on Various Query Optimization Technique, " International Journal of Computer Science and Mobile Computing, Vol. 5 Issue 8, August 2016.

[8] Imran Ashraf, Tayyaba Altaf, Ayesha Rashid, "Cloud Data Storage and Query Procesing System : A Review," International Journal of Scientific and Engineering Research, Vol. 5 Issue 6, August 2014.

Land Registry using Blockchain

Krati Paliwal,Siddhi Poojari,Mayur Singal,Sandip Bankar
Information Technology
Thakur College of Engineering Technology
Mumbai, India
kratip11@gmail.com,siddhipoojari7@gmail.com,mayursingal9@gmail.com,sandipbankar1@gmail.com

Abstract - **A blockchain is a growing list of records, called blocks, that are linked using cryptography. Blockchain contains a comprehensive approach in industries which requires to be protected from corruption, human error or human intervention. Land Registry is one of the use cases that involve plenty of intermediaries to place trust within the system. The present solutions in place are out of date. Land registration generally describes systems by which matters concerning ownership, possession or other rights in land will be recorded (usually with a bureau or department) to produce evidence of title, facilitate transactions and to prevent unlawful disposal. The knowledge recorded and therefore the protection provided will vary by jurisdiction .The current land registry system is rife with corruption and inefficiencies. With blockchain, we've an opportunity to mend many of those problems.**

I. INTRODUCTION

Land registration is a process by which matters concerning ownership, possession or other rights pertaining to the respective land may be recorded (usually with a office or department) to supply proof of ownership, facilitate transactions and to avoid discrepancy. Keeping track of who owns which pieces of land or property is difficult once you have hundreds or maybe thousands of years of land records and therefore the current systems in situ are outdated.

It's common to encounter discrepancies within the paperwork, including forged documents, counterfeit titles and, in some cases**,** a whole loss of all documentation. Such situations result in expensive court battles between conflicted parties. The transparent nature of blockchain can make it possible to trace how property changes hands. Blockchain's immutable, auditable and traceable features are enticing governments around the world to implement the decentralized technology within the land registry process. Using the blockchain technology it's possible to make a land registry and a history of transactions which will be easily verified at any given point. [1]

II. OBJECTIVE AND SCOPE

The objectives our project aims to achieve and its scope are:

1. Develop an accessible database to store land titles and transactions
2. To provide high capacity and throughput for millions of records
3. Data immutability that brings trust and auditability to the records
4. Quick retrieval of transaction histories

III. PROBLEM DEFINITION

Currently no solution exists to the issues faced within the existing land registry system. Thus, our system aims at building a true time record maintenance system that captures various attributes of land by using blockchain because the key technology. The system will let users to shop for and sell land securely while maintaining authenticity of the ownership. We've divided the project into 2 stages.

Stage I - will focus on shaping a base for the project. With intensive arranging, examination and structure, we'll guarantee that the important usage is smoother. Planning aims at deciding the agile methodology for arranging our land registry project. Analysis focuses on preparing an analysis of our own project. founded an itemized investigation on present enlistment and approval forms and conquering its impediments and execution

benchmarks. the planning phase focuses on designing of event system application.

Stage II will concentrate on building up the undertaking and guaranteeing that the planned highlights will function as we needed them to. This stage includes developing phase where,we will send a Blockchain in which the qualifications of purchasers and venders as hashes is put away likewise in this stage we will build up an interface for our users to list our their properties and get them approved. Testing includes doing the different test framework application. The last phase is deployment where beta testing for distinguishing any further mistakes, bugs and enhancements that can be performed.

IV. TOOLS & TECHNOLOGIES

Etherium: Ethereum is a distributed public blockchain network. Ethereum specializes in running the programming code of any decentralized software. In the Ethereum blockchain miners work to earn Ether, a kind of crypto token that fuels the community. Beyond a tradeable cryptocurrency, Ether is also used by application builders to pay for transaction expenses and offerings on the Ethereum network. There is a 2nd type of token this is used to pay miners fees for which includes transactions in their block, it is known as gasoline, and each clever contract execution requires a certain quantity of gasoline to be despatched at the side of it to trap miners to position it inside the blockchain. [5]

Metamask: MetaMask is a bridge that allows you to connect the distributed web with a normal web browser. It allows you to run Ethereum based Dapps right in your browser without implementing a full Ethereum node. It is a web browser extension that injects the Ethereum web3 API into every website's javascript context, so that Dapps can read from the blockchain.

MetaMask also lets the user create and manage their own identities, so when a Dapp wants to perform a transaction and write to the blockchain, the user gets a secure interface to review the transaction, before approving or rejecting it. Since it adds functionality to the conventional browser context, MetaMask requires the permission to read and write to any webpage.

Solidity: Solidity is an object-orientated language particularly evolved for contract writing. It is a high-stage language, which inherits traits from C++, Python, and JavaScript. The Solidity compiler compiles your source code into bytecode that runs on Ethereum Virtual Machine.Solidity is a statically-typed programming language designed for developing smart contracts that run at the EVM. With Solidity, builders are able to write programs that put into effect self-enforcing commercial enterprise logic embodied in clever contracts, leaving a non-repudiable and authoritative record of transactions. [5]

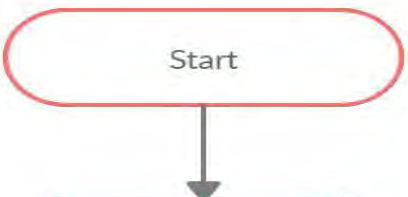
Ganache: Ganache could be a personal blockchain for Ethereum development you'll use to deploy contracts, develop your applications, and run tests. It's available as both a desktop application in addition as a command-line tool (formerly referred to as the TestRPC).

Ganache Accounts shows the accounts generated and their balances. Blocks shows each block as mined on the blockchain, together with gas used and transactions. Transactions lists all transactions run against the blockchain. Contracts lists the contracts contained in your workspace's Truffle projects. Events lists all events that are triggered since this workspace's creation. Ganache will try and decode events triggered by contracts in your Truffle project. Logs shows the logs for the server, which is beneficial for debugging. [6]

Web3: Web3 is a extensive movement and an inclusive set of protocols aiming to make the web and the net extra decentralized, verifiable, and relaxed. WEB3 is the imaginative and prescient of the serverless internet, the decentralized web. An net in which users are in control in their very own statistics, identity and future.

Remix IDE: Remix is a Solidity IDE that’s used to write, compile and debug Solidity code.
Solidity is a high-level, contract-oriented programming language for writing smart contracts. It was influenced by popular languages such as C++, Python and JavaScript.

V. METHODOLOGY



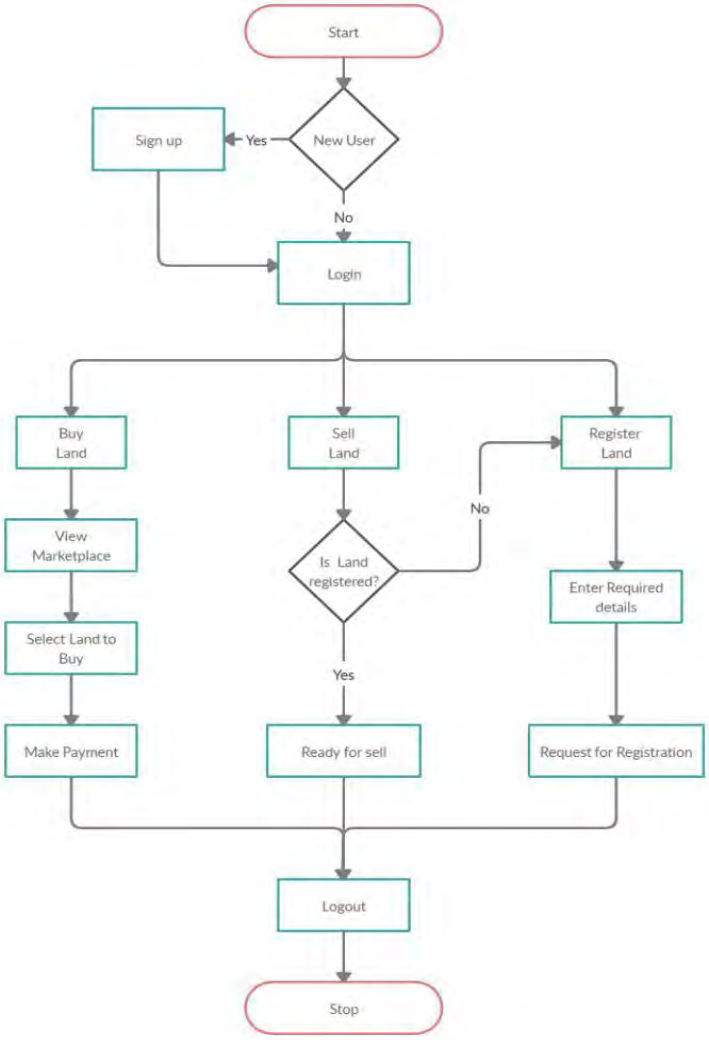


Fig 5.1: Flow chart: User [3]

Fig 5.2: Flow chart: Registrar [3]

User: For buying land

1. The user will register on the portal.
2. The user can view the marketplace and check the different properties listed.
3. As per his requirements the user can select the land to buy.
4. On selection of land, the payment can be made.
5. When user buys property, registration need not be requested as it will already be a part of blockchain from the seller's end.

User: For selling land

1. The user will register on the portal.
2. For registered property, the user can list it out for sale.
3. If the property is not registered, the user will have to fill in the required details and documents and request for registration,
4. Upon completion of registration process, the property can then be listed for sale.

Registrar

1. The registrar will login into the portal.
2. He can view the registration requests on the dashboard.
3. The registrar can accept and reject requests based on the details uploaded by the requester.
4. The accepted requests will be stored as a block in the blockchain.

VI. FUTURE SCOPE

Currently the proposed system can be used to authenticate and validate property ownership and allow the user to put up properties for sale.as the verified user ownership information gets stored as a block in blockchain, it becomes secure due to security aspects of blockchain. The information will be available to different departments as per their usage. The resultant system will serve as an easy, quick and cost efficient record maintenance and retrieval system. The current project can be extended by increasing the area coverage. The proposed system can be extended to different states. The project can be further integrated with maps for better geographical identification and understanding. The system can be used as a secure and trustworthy resource to store, maintain and retrieve property information of various users over a given region.

VII. SUMMARY

The project aims at delivering a trustworthy and reliable system that will make land registration process hassle-free while maintaining security. Large amounts of records can be stored easily. The transparent nature of blockchain can make it possible to trace how property changes hands. Also, the challenges of time delays, involvement of middlemen and frauds can be overcome.

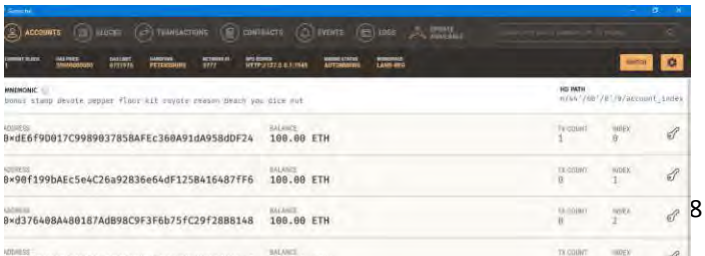
The capacities required for the venture are reasonably in the attainable range. A GUI is required, for the clients to make their records and rundown out their properties for selling or for enlisting another one. The enrollment office requires an interface for confirming the ownership. Verified data will be included to the blockchain. The GUI will be kept basic and moderate. This application will have numerous positive applications on society. It will improve the validity of the clients and give validness to proprietorship. The philosophy that we decided for our task is Agile. Light-footed enables us to be adaptable, oversee evolving necessities, deal with the regularly expanding extension just as get purchaser viewpoint

.

VIII. RESULTS

Fig 8.1: Land Registry Smart Contract

Fig 8.2: Land Registry Smart Contract [6]



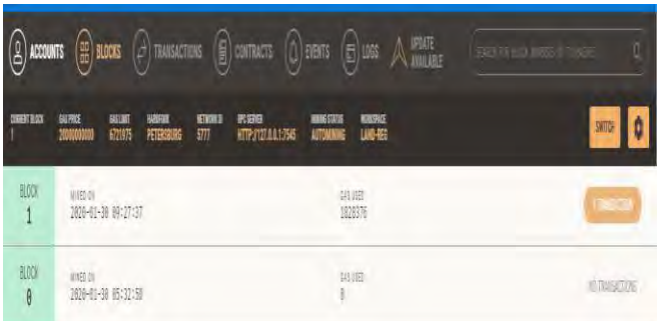


Fig 8.3: Mined Blocks

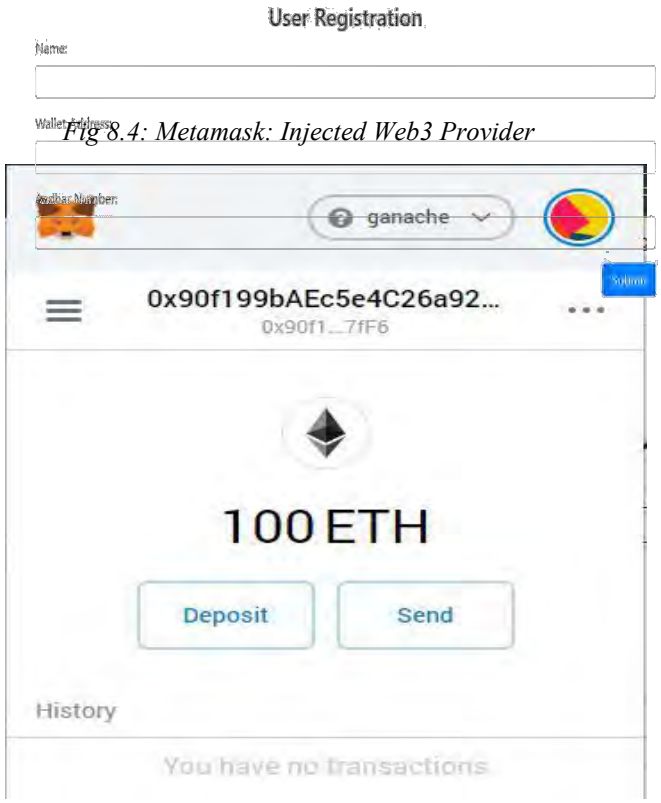


Fig 8.4: Metamask: Injected Web3 Provider

Fig 8.5: User Registration Page

IX. CONCLUSION

The Blockchain based Land registry system will facilitate integration and validation of land records over a given region. The system allows users to buy, sell and register new properties. It is easy to use and a secure way of buying and selling land. Blockchain provides the key feature of security which allows the registrar to validate the user's authenticity in a hassle freeway. As the verified user's record is entered in a block in blockchain its can surely be relied upon. The problem of discrepancies and false records is eliminated in our system. Thus, it can be guaranteed that the property listed by a user belongs to him only. This makes the system trustworthy.

References

fflesuite/ganache

[1] Valentina Gatteschi, Fabrizio Lambert, Claudio Demartini, Chiara Pranteda,Víctor Santamaría, To Blockchain or Not to Blockchain:That is the Question, 2018

[2] Vitalik Buterin, Ethereum: A next generation smart contract & decentralised application platform, 2014

[3] Alex Mizrahi, A blockchain based property ownership recording system, 2015

[4] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, Hawk:The blockchain model of cryptography and privacy-preserving smart contracts, 2017

[5] Maximilian Wohrer , Uwe Zdun : Smart contracts: security patterns in the ethereum ecosystem and solidity, 2018

[6] Ganache Tool. [Online]. Available: https:// github.com/tru

Voting Using Blockchain Technology

Dr Shabnam Sayyad Dr Farook Sayyad[2]
[1] Department of Computer Engineering, AISSMS College of Engineering,
[2]Department of Mechanical Engineering , Dr D Y Patil School of Engineering, Lohegaon,
Pune, India
ssshaikh@aissmscoe.com,fbsayyad@gmail.com

Abstract

The increasing list of records that are widely known as blocks, which are linked using the cryptographic hash of the previous blocks, is referred to as blockchain. The Blockchain technology allows distributed public records that hold absolute data in a secure and encrypted way and ensure that transactions can never be changed. The Bitcoin and other crypto currencies are the most prevalent instances of blockchain practices.

Now-a-days there are many researches in the blockchain technology are used to find solutions for so as to find real-world problems.

After Bitcoin, many digital currencies incepted. For many years it was misunderstood that blockchain can be used only for maintaining the online currency transaction record. There was less awareness about the applications of blockchain

One of the most appropriate instance is in inheriting features of the Blockchain for online Voting or decentralized Voting. In this paper, we propose a voting solution that is made using the Ethereum protocol. It utilizes the properties of smart contracts to impose strict rules surrounding the ballots of an election. These ballots are both autonomously and collectively provable and maintain all of the desired properties of the blockchain (like immutability). This is achieved by maintaining the voter privacy and ballot integrity. The system described shows clear potential for Blockchain technology to become a dominant part of applications coveting to provide transparency and security in public states.

Keywords—Blockchain, Online Voting, Ethereal, Encryption, Blockchain voting features.

I. INTRODUCTION

In Modern technology Online, voting is thought as the most glitchy because of its major security issues. Online voting is only beginning to make headway in modern society. It might seem strange that the acceptance of this method of voting has been so slow, but if you take a closer look at the issues involved, the reason will become clear. Online or remote voting iscertainly convenient, but it also leads to vast chances for spoofing the results. A conceded voting protocol can lead to large-scale changes to the choices made by voters. This means that online voting imposes extremely severe necessities on the security of every aspect of voting. We believe that the blockchain technology is the omitted link in the architecture of a practicable online voting system. It's been challenging to develop an online voting system to satisfy the legal requirements of a democratic country. Blockchain technology has come up with interesting features, which can make the impossible looking process a easy victory. Blockchain offers an infinite range of applications among which online voting is the one. The diverse features of Blockchain technology can make a positive change Blockchain technology offers the diverse features which can make the change. With rise in the population increases to maintain democracy is difficult, trying to satisfy all strata of people. Using blockchain technology uncertain issues related to people's identity and about transparency in long distances and complex trades, like in election systems can be maintained. This problem can be solved by building a more transparent system through which people can vote. blockchain technology that can support us reduce our uncertainties about identity. In this paper, Our implementation is to be deployed on Ethereum's testnet to demonstrate usability, scalability and efficiency by making blockchain based voting system.

II. DEFINITION

The traditional paper ballot voting can be replaced by voting using Blockchain technology. The security issues can be easily dealt as I t is cryptographically secured, transparent, public ledger technology which can eliminates the trust breach between central authority and voters, and hence providing the portable and democratic voting solution with only a minor

requirement of an individual knowing to operate a mobile phone.

A blockchain is a "cryptographically secure transactional singleton machine with shared-state." "Cryptographically secure" means that the creation of digital currency is secured by complex mathematical algorithms that are obscenely hard to break. "Transactional singleton machine" means that there is a single canonical instance of the machine responsible for all the transactions being created in the system. In other words, there's a single global truth that everyone believes in. "With shared-state" means that the state stored on this machine is shared and open to everyone.[6].

The blockchain is built on a decentralized network, which retains the transactional records in the form of blocks in chronological order. The first block is identified as a genesis block. The other succeeding blocks are linked to the previous block. Hash value of the previous block is stored in every block thus linking all the blocks together. Hash is generated depending on the transaction records in the block. Changing the record in block changes the hash and thus makes the chain invalid. This makes the Blockchain immutable.

III. LITERATURE REVIEW

Author implemented the system in which a voter is given a token that allows them to cast vote for their chosen candidates [2]. Privacy is a seminal work that has research on end-to-end (E2E) Evoting which has become a flourishing field. Familiarly, the notion of being E2E verifiable refers to have two attributes: First, every voter can verify if their vote has been cast as desired, recorded as cast. Second, anyone can verify if all votes are tallied as recorded[5]. The proposed number of E2E verifiable schemes and some of these are used in practice[6]. A voting system, called DRE-i (DRE with integrity) [15], to achieve E2E verifiability without involving any tallying authorities (TAs)[1]. DRE-I has precipitation strategy that requires the pre-computed data to be securely stored and accessed during the voting phase. This introduces the possibility for an rival to break into the secure storage module and com- promise the privacy of all ballots. To overcome this disadvantage, they provided the voting system DRE-ip(DRE-i with enhanced privacy). DRE-ip achieves E2E verifiability without TAs and concurrently a significant stronger privacy guarantee than DRE-i [3]. The two assumptions in which the two probability distribution (g a , gb , gab): a, b are randomly and independently chosen from Z q and (g a , gb , gc): a, b, c are randomly and independently chosen from Z q are computationally indistinguishable in the security parameter n = log(q). Our Proposed System works over an elliptic curve in an ECDSA like group setting or a DSA like multiplicative cyclic group setting where the decisional Diffie-Hellman (DDH) assumption is implemented. [5]

Ethereum is an open-source, public, Blockchain based distributed computing platform and operating system featuring smart contract (scripting) functionality. Blockchain voting can be implemented in different phases[4]. Most common stages are as follows: Initialization, Registration, Ballot casting, verification, Tallying results, Revealing results.

Various systems use a different methods by combining a few phases together. The proposed system is using the following phases,

1) Voter registration is the phase in which voter provides personal information. The information is verified by authorized personnel and stored into the blockchain.

2) Creating private Blockchain for voter registration information is the step in which a new Blockchain combined with Merkle hash tree is created to store the voter's information Merkle hash tree is also known as a binary hash tree. Merkle hash tree is used to store a large set of data and to efficiently summarize and

verify the integrity of that large data set. Merkle tree has the following features: i) Ability to verify whether a transaction is included in a block ii) Light-clients. iii) Overall performance and scalability iv) Simplified Payment Verification or (SPV).

3) Voter Authentication is the step in which the physical verification of voter is done based on the record stored in private Blockchain.

4) Voting and tallying phase is the last step where actual voting is done and the results are tallied and displayed to the voter.

Features of Blockchain voting:

Eligibility: This property states that only eligible users can vote. Eligibility criteria can be different for different countries. One of the criteria is the age of the user who is casting the vote. In India person who is above 18 yrs old is allowed to cast the vote.

Privacy: Privacy is one of the most important aspects of democratic voting. Voters privacy should be maintained. No one should be able to know how a particular person voted or to whom the particular voter voted.

Coercion resistance: No one should be able to force the voter and should not have the ability to distinguish between whether the voter voted the same way he/she was instructed to vote.

Physical verification: Physical verification is necessary to know whether the actual voter voted or the bogus person voted on the identity of another voter.

Forgiveness: The voter should be able to alter the vote at any time before the election ends. This is related to the coercion resistance. Even if the coercer forced the voter, a voter should be able to change the vote afterward.

Verifiability: This property states that everyone involved in the voting process should be able to verify the results. This brings transparency in the election. Also, an individual voter should be able to verify whether his/her vote is counted or not.

Immutability: The voter's vote should be immutable. No one should be able to change the vote of any voter without proper concern of the voter. All the records should be immutable.

IV. IMPLEMENTATION OF BLOCKCHAIN VOTING

There are many platforms to implement Blockchain voting but the most used platform is Ethereum. It is open-source, public, Blockchain based distributed computing platform and operating system featuring smart contract (scripting) functionality. When using Ethereum, computational expenses are manifested as gas price. An initial gas value is set, which can be used to perform the operations. Blockchain voting can be implemented in various phases. Most common phases are as follows: Initialization, Registration, Ballot casting, verification, Tallying results, Revealing results.

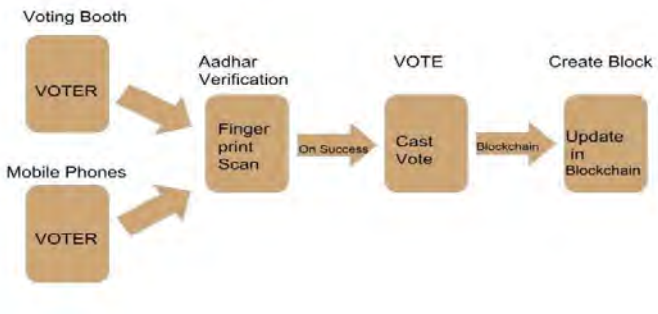


Fig.1 Architecture diagram of the system

We will concern about the features achieved by implementing Blockchain voting this way. Since our voting is based on Blockchain, many features get inherited from the Blockchain itself, such as End-to-end verifiability, transparency, immutability etc. This

feature of Blockchain solves some of the legal requirements of democratic voting.

But with this advancement there comes some possible threats such as maintaining anonymity becomes harder. Different research papers have proposed various implementations to solve these problems. Those implementations are as follows.

Shamir's secret sharing scheme is used by polys voting system. The idea is based on the concept that one will need k points to interpolate the polynomial of degree k-1. for e.g, 2 points are must for drawing a line segment or 3 points are must for drawing a curve. So one can share a secret among n person in such a way that it can only be recovered by k people(k<=n), we need to hide that secret in the formula of a polynomial of degree k-1. This implementation maintains the privacy of every user even if the data is visible to everyone. So the problem of transparency and privacy is solved at the same time.

Homomorphic encryption is another concept which can be used to provide anonymity. Homomorphic encryption is the encryption technology which can be implemented using the exponential ElGamal cryptosystem. The basic idea is to allow the authorities to tally the ballots without actually decrypting them thus providing the privacy and security to the system. Exponential ElGamal is a cryptosystem which is used for encryption of the voter data and provides the additive Homomorphic property to the system.

Another way of implementing homomorphic encryption is Pallier encryption. Homomorphic encryption is nothing but to perform computation on encrypted data such that computations are performed originally on decrypted data. Doing a fully modular form of multiplication in fully homomorphic encryption is computationally intensive and slow. So, partial homomorphic encryption scheme such as pallier encryption is used. This adds two ciphertexts and multiplies one ciphertext with another plain ciphertext.

V. COMPARISON OF RESULTS

One-time Ring Signature techniques can also be used to protect anonymity. Un-linkable signature is used at the receiver end. In this scheme, no observer can determine whether the transactions are sent to a particular receiver address or the two addresses are associated with each other. To keep the anonymity at sender side ring signature technique is used. Here the sender generates a unique ring structure with a fracture in it using the public key produced by other signers and completes the ring using the private key.

Error! Reference source not found. shows results of experimentation.

	Polys[5]	Bronco Vote[1]	Ranked Choice Voting[2]	Bit Congress[4]	Follow My Vote[4]
Eligibility	yes	yes	yes	no	yes
Anonymity	yes	yes	yes	yes	yes
Verifiability	yes	yes	yes	yes	yes
Integrity	yes	yes	yes	yes	yes
Physical Verification	-	no	no	-	-
Forgiveness	-	-	yes	yes	yes

Even though Blockchain provides many features, there are some limitations and issues of Blockchain. The first issue is of majority attack, which is nothing but, if someone has more than 51% of computational power then he/she can modify the transaction data. Another issue related

Blockchain is Fork Problem. When the system comes to a new agreement or new version, then the Blockchain network is divided into two types, new nodes, and old nodes, so after version change, old

nodes could not agree with the new nodes and thus a problem occurs.

VI. CONCLUSION

In this paper, we have presented the features of the online voting system and how different systems have addressed those features. Many systems have managed to address most of the features effectively. Some of the features are inherited from the blockchain and remaining are solved by some encryption techniques like Homomorphic encryption. Although most of the systems have effectively used the features of Blockchain, there are some features of online voting which are not addressed yet.

References

[1] Gaby G. Dagher, Praneeth Babu Marella, Matea Milojkovic, Jordan Mohler, "BroncoVote: Secure Voting System using Ethereum's Blockchain" , ICISSP 2018 - 4th International Conference on Information Systems Security and Privacy, 2018

[2] Xuechao Yang, Xun Yi, Surya Nepal, Andrei Kelarev, and Fengling Han, "A Secure Verifiable Ranked Choice Online Voting System Based on Homomorphic Encryption", DOI 10.1109/ACCESS.2018.2817518, 2018

[3] Baocheng Wanga, Jiawei Suna, Yunhua Hea, Dandan Panga, Ningxiao Lua, "Large-scale Election Based On Blockchain", International Conference on Identification, Information and Knowledge in the Internet of Things, Available online at www.sciencedirect.com, 2017

[4] Freya Sheer Hardwick, Apostolos Gioulis, Raja Naeem Akram, and Konstantinos Markantonakis, "E-Voting with Blockchain: An E-Voting Protocol with Decentralisation and Voter Privacy",2018

[5] Polys - Online voting system, Whitepaper. [online] Available : https://polys.me/assets/docs/Polys_whitepaper.pdf

[6] Preethi Kasireddy, "How does ethereum work anyway", [online] Available : https://medium.com/@preethikasireddy/how-does-ethereum-work-anyway-22d1df506369

[7] Andrew, "Blockchain Fundamentals #1: What is a Merkle Tree?", [online],Available: https://medium.com/byzantine-studio/blockchain-fundamentals-what-i s-a-merkle-tree-d44c529391d7

[8] Somnath Panja, Bimal Kumar Roy, "A secure end-to-end verifiable e-voting system using zero knowledge based blockchain", [2018]

[9] Silvia Bartolucci, Pauline Bernat, Daniel Joseph, "SHARVOT: secret SHARe-based VOTing on the blockchain", ACM/IEEE 1st International Workshop on Emerging Trends in Software Engineering for Blockchain, [2018]

[10] Iuon-Chang Lin and Tzu-Chun Liao, "A Survey of Blockchain Security Issues and Challenges", International Journal of Network Security, Vol.19, No.5, PP.653-659, [2017]

[11] Shaan Ray, "Blockchain Forks" [online] Available: https://hackernoon.com/blockchain-forks-b0dca84db0b0

[12] Lionel Dricot and Olivier Pereira, "SoK: Uncentralisable Ledgers and their Impact on Voting Systems"

Performance evaluation of different stress factors for Type-2 Diabetes Mellitus using logistic regression model

Rohini Patil[1], Dr. Kamal Shah[2]
Terna Engineering College, Navi Mumbai[1]
Thakur college of Engineering and Technology Mumbai.[2]
rohinipatil@ternaengg.ac.in, kamal.shah@thakureducation.org

Abstract — **Rising prevalence of diabetes is a concern today. Not only in India but across the world. Although awareness has increased to some extent, many people with diabetes have limited knowledge about risk factors before the diagnosis of disease. For chronic disease prevention there is a necessity to find out such risk factors and manage them appropriately. Statistical techniques can be employed to understand the risk of diabetes in different age group of people. This research is planned to predict risk of type 2 diabetes in people with different age groups according to different levels of stress factor and to find risk of diabetic. The Proposed system works on six different stress factors.To develop an adaptive model the preprocessing step has been applied along with logistic regression. The results suggest that the logistic regression algorithm may be useful for predicting the risk of diabetes among non-diabetic people.**

Keywords—Diabetes Mellitus,risk,Stress

I. INTRODUCTION

Machine learning algorithms can build a model from data used for prediction, decision making or solving some task. Predictive analytics uses statistical, analytical queries or machine learning methods for predicting potential outcomes [1]. Past and current data can be used for predicting future activity, behaviour and trends. Healthcare is one of the areas where predictive analytics can be of significant value. Today, diabetes represents a significant health burden among families and in turn on the country. The incidence and prevalence of both are on rise worldwide and in India.

Type 2 diabetes mellitus is a chronic condition which occurs because of decreased synthesis of insulin by the pancreas or ineffective utilization of insulin. Diabetes mellitus is broadly divided into three main types; type 1 (insulin dependent) diabetes, type 2 diabetes and gestational diabetes.

In type 1 diabetes, because of the deficiency of insulin, body does not properly metabolize food which results in high blood glucose level. Because of inability of pancreas to produce insulin it has to be given from outside hence patients need lifelong injections of insulin. Type 2 diabetes occurs due to ineffective utilization of insulin [2]. Type 2 diabetes mellitus is common than insulin dependent. Type 2 diabetes can be prevented by healthy life- style.

Gestational diabetes occurs during pregnancy [3]. Women with gestational diabetes have high risk of developing type-2 diabetes. Gestational diabetes is diagnosed regular checking of blood glucose during pregnancy.

In pre-diabetes, blood glucose levels are elevated but not as high as diabetes mellitus [4]. Pre-diabetes is a risk factor for diabetes so precaution need to be taken at earlier stage.

The two well-known international associations i.e. World Health Organization (WHO) and International Diabetes Federation (IDF) produce report on epidemiology. According to the report of the World Health Organization (WHO) published in 2016, diabetes is the seventh leading cause of death [5].

According to the International Diabetes Federation (IDF), India **is** the second largest country with diabetes cases with an estimated 73 million adults between the age group of 20-79 years. Current burden of diabetes in India is 49% of world burden and this is expected to reach 151 million by 2045, posing a major health challenge for the country [6]. Diabetes is associated with several complications including those related to eye, nerves, arteries, kidney and heart [7].

This research is carried out to predict type 2 diabetes risk based on daily lifestyle activities such as different stress factors and other indicators used to predict the disease which creates a simple diagnostic application. This simple application may be useful for diagnostic purpose and the results may also be useful for improving health.

II. REVIEW OF LITERATURE

For predicting risk of diabetes among people with different age groups based on life style of a person data mining approach is used. Decision trees model is used for predictive analysis. The results suggested that the people whose age is above 45 years have more prevalence of diabetes and the risk factors included hypertension, lipid level, body mass index (BMI) and the history of myocardial infarction [4].

Another group of researchers used traditional and hybrid approach as a predictive model. They showed hybrid model to be more accurate than traditional model [8]. In another model diabetic patients were classified in three age groups i.e. 18 to35, 36 to55 and >55 years, using adaboost and decision tree model. The decision tree model they have used is J48 [9].

For accuracy improvement researcher developed a model with two level algorithm i.e. unsupervised k-means and logistic regression applied on 14 attributes for predicting the risk of type 2 diabetes mellitus [10].

The researcher applied Chi-squared test and used classification and regression tree (CART) model has

been shown to provide accuracy of 75% for the prediction of diabetic based on daily lifestyle. Blood Pressure was identified as a significant factor for the development of diabetes along with junk food, sleep, family history, rice intake and the physical activities performed [11].

For prediction of risk of diabetes investigators develops a web application by using classifiers and a real data set. They have been used Different models like Artificial Neural Network, Decision Tree, Naïve Bayes, Logistic Regression, and Random Forest algorithm for prediction analysis. Out of these, Random Forest algorithm was selected for the prediction of diabetes risk as it gives best performance. In this approach, input variables are randomly selected and used for creating the application [12].

Harris et al. shows how perceived stress is a strong risk factor for type 2 diabetes. They show Direct and indirect effect of stress on diabetes through risk factors, like BMI, hypertension, and physical activity. This study is only female centric [13].

Martinez et al. shows how stress affect diabetes. They found an increased risk for T2D in people exposed to stressful conditions, traumatic events, mental health problems, depression. [14].

Perveen S. et al. shows the relationship between diabetes mellitus and individual risk factors of Metabolic Syndrome, Logistic regression is used for validation of predictive power of risk along with data sampling techniques [15].

III. MODEL AND ALGORITHM

In this study, we collected the data through paper based and web-based questionnaires. Although different factors such as gender, history of diabetes, daily activities, amount of sugar, exercise, stress, food habit, BMI, family history, systolic blood pressure, blood glucose, diastolic blood pressure and smoking playing a key role in the development of risk of type 2 diabetes, we focused only on stress related factors such as, Workload, poor salary, deadlines, travel, repetitive work, unplanned work for prediction of type 2 diabetes.

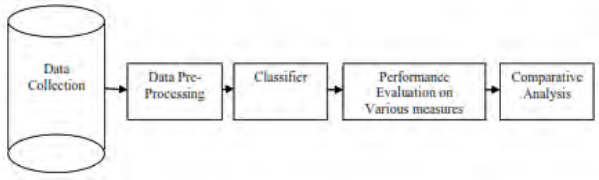


Figure 1. Proposed system model

Figure 1 above shows proposed system model used in our study. Data were collected from diabetic and non-diabetic people both. The collected information was entered in the excel sheet, cleaned for homogeneity means firstly. For building a prediction model data analysis was performed by applying preprocessing step and a suitable machine learning model which improves the accuracy.

A. Data preprocessing

In this cross-sectional study (single visit), data were collected from225 people including both diabetic and non-diabetic people. Out of 225 people, 207 responses were eligible for analysis. Other 18 responses were not considered for data analysis because of the incomplete entries or some missing information.

Some variables Body mass index (BMI) and height required a transformation. Height was collected in feet and then converted to (centimeters) whereas BMI was calculated by dividing weight (kg) by height (meters) squared. The questions were related to stress related factors such as workload, poor salary, deadlines, travel, repetitive work, unplanned work, for prediction of type 2 diabetes. Table 1.describes different attribute used in dataset.

Sample variables

TABLE 1: ATTRIBUTE DESCRIPTION

No.	Variables	Description	Possible value
1	Sex	Sex	1: Male 2: Female
2	Age	Age(years)	1 :18 to 34 years 2: 35 to 44 years 3: 45 to 54 years 4: 55 to 64 years 5: 65 above
3	Weight	Weight(kg)	Numeric
4	Height	Height(m)	Numeric
5	BMI	Body Mass Index (Kg/m^2)	1:underweight 2:eutrophic 3:overweight 4: obese
6	Stress	Stress	1: Very low 2: Low 3: Average 4: High 5: Very high

B. Logistic regression model

which can map input data item to a given category. A statistical approach for analyzing a dataset where one or more independent variables which determine an outcome using logit function. It is binary classification problem. The outcome is measured in the form of dichotomous variable means whether a person is diabetic or non-diabetic. Logistic regression algorithm is used for predicting and classifying medical problem, for automatic disease diagnosis.

Logistic regression algorithm, called as logit model. It uses linear equation with independent predictors. Consider n input variables where their values are indicated by x1, x2, ..., xn. Let z be a probability to indicate an event will occur and 1-z be a probability such that event will not occur. The logit model shown in the following equation,

$$Z = \frac{e^y}{1+e^y} \quad (1)$$

$$\log\left(\frac{z}{1-z}\right) = y \quad (2)$$

$$\log\left(\frac{z}{1-z}\right) = \text{logit}(z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (3)$$

Where β0 is the intercept and β1, β2… βn are the regression
coefficient

IV. EXPERIMENTAL RESULT

Using python toolkit, we analysed and evaluated our model using following aspects .
K-fold cross validation

Cross -validation is a technique of reducing overfitting. In our study we have applied 10- fold cross validation. Model was
trained and tested 10 times. In this initial samples were partitioned into 10 sub-samples from that 9 subsamples were used for training and each subsample was retained as validation data.
Performance parameter
To assess the performances of models, Let TP as true positives (correctly predicted number of instances), FP as false positives (incorrectly predicted no of instances), TN as true negative (the no of instances are negative but predicted correctly), FN as false negatives (the no of instances are negative but predicted incorrectly). From the outcome of accuracy some substantial performance parameters such as precision Recall, F1-score are as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} X100 \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} x100 \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} X\,100 \quad (6)$$

$$\text{F1-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Table 2.shows Statistical measure of various stress related
factors.
Table 2.Accuracy parameter for different stress related factor

Factor	Parameters		
	Precision	*Recall*	*F1-score*
Workload	0.75	0.76	0.73
Poor Salary	0.70	0.73	0.70
Deadlines	0.75	0.76	0.73
Travel	0.77	0.78	0.76
Repetitive work	0.75	0.76	0.73
unplanned work	0.75	0.76	0.73

Figure 2.shows comparison of different stress factors such as workload, poor salary, deadlines, travel, repetitive work and unplanned work.

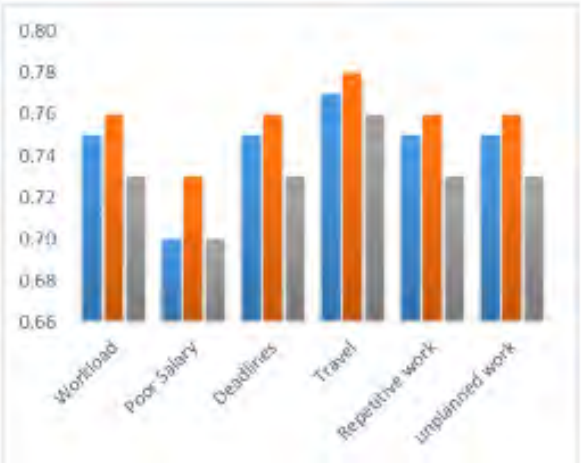


Figure 2.Statistical comparison on various measures

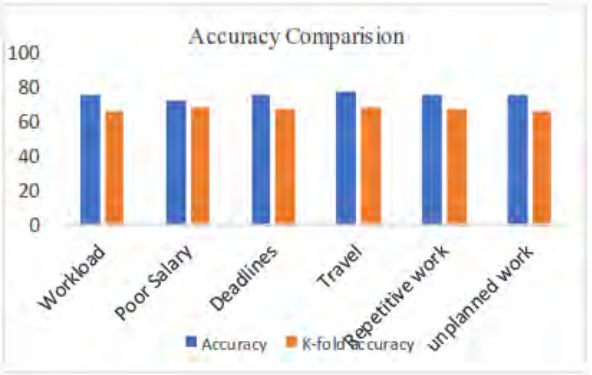


Figure 3. Accuracy comparison of stress parameter
For chronic diseases, stress must be reduced. Comparison of different stress parameters is shown in figure 3.

V. LIMITATIONS

Small number of responses, limited number of risk factors and use of single analytical method for prediction modeling are the limitations of our study. These findings need to be evaluated by using larger data sets of diabetes and non-diabetes people with multiple prediction modeling techniques. Considering the above listed limitations readers should carefully extrapolate findings of our study

VI. CONCLUSION AND FUTURE WORK

The parameter such as gender, age, body mass index (BMI) along with the Stress factors such as Workload, poor salary, deadlines, travel, repetitive work, unplanned

work. Stress is identified as a important risk factor for development of diabetes, in this work, use of logistic regression model provided accuracy of 76%, for prediction of diabetes. Further research can be done to evaluate effect of gender, age and other risk factors on the risk of diabetes

References

[1] Brown DE, et al. Predictive analytics. Washington: IEEE Computer Society; 2015

[2] Sanakal R, Jayakumari T. Prognosis of diabetes using data mining approach-fuzzy C means clustering and support vector machine. Int J Comput Trends Technol. 2014;11(2):94–8.

[3] Lakshmi KR, Kumar SP. Utilization of data mining techniques for prediction of diabetes disease survivability. Int J Sci Eng Res. 2013;4(6):933–40.

[4] Repalli P. Prediction on diabetes using data mining approach. Stillwater: Oklahoma State University; 2011.

[5] http://www.who.int/en/news-room/fact-sheets/detail/diabetes accessed on 18th nov.2018

[6] https://www.idf.org/our-network/regions-members/south-east-asia/.../94india.html assessed on 18th Nov.2018

[7] Motka R, et al. Diabetes mellitus forecast using different data mining techniques. In: Computer and communication technology (ICCCT), IEEE, 4th international conference. New York: IEEE; 2013.

[8] Jayanthi N. , B. Babul V. Rao S.N.Survey on clinical prediction models for diabetes prediction. J Big Data 2017; 4:26 DOI 10.1186/s40537-0170082-7

[9] Sajida Perveena, Muhammad Shahbaza, Aziz Guergachib, Karim KeshavjeecPerformance Analysis of Data Mining Classification Techniques to Predict Diabetes Procedia Computer Science 82 (2016) 115 – 121

[10] Wu H,YangS,HuangZ,HeJ,WangX.Type 2 diabetes mellitus prediction model based on data mining.Informatics in medicine unlocked 2018;10:100-107

[11] Aysh anand, Divya shakti Prediction of diabetes based on personal lifestyle indicators,In :1st international conference on next generation computing technology:IEEE;2015

[12] NongyaoNai-aruna,,RungruttikarnMoungmaia,Comparison of Classifiers for the Risk of Diabetes Prediction, Procedia Computer Science 69 (2015) 132 – 142

[13] Harris ML,Oldmeadow C, Hure A, Luu J, Loxton D, Attia J. Stress increases the risk of type 2 diabetes onset in women: A 12-year longitudinal study using causal modelling. PLoS ONE 12(2):e0172126.doi:10.1371/journal. pone.0172126

[14] Martinez A, Sanchez W, Benitez R, Gonzalez Y, Mejia M, Otiz J. A job stress predictive model evaluation through classifier's algorithms. IEEE Latin America Transactions, VOL. 16, NO. 1, JAN. 2018

[15] Perveen S.Shahbaz M. Keshavjee K. Guergachi A. Metabolic Syndrome and Development of Diabetes Mellitus: Predictive Modeling Based on Machine Learning Techniques DOI 10.1109/ACCESS.2018.2884249, IEEE Access

[16] Min Chen, Yixue Hao, Kai Hwang, Fellow, IEEE, Lu Wang, and Lin Wang Disease Prediction by Machine Learning over Big Data from Healthcare Communities IEEE access 2016; 2169-3536

Healthcare Application using Artificial Intelligence

Mitesha Jadhav[1], Ahrar Khan[1], Dipesh Gaggar[1], Aaditya Desai[2]

[1](Information Technology, Thakur College of Engineering and Technology/ Mumbai University, India)
[2](Assistant Professor, Information Technology, Thakur College of Engineering and Technology/ Mumbai University, Indi

Abstract— **While Developing a healthcare application Managing Patient's personal data has always been a challenge since the app requires information such as symptoms, Date of birth and his/her current health state. In the IT Sector, the integrity and Confidentiality is the major concern for every organization. As the rise in the attacks on the data are at fire nowadays, The Uphill task was to make the app more secure and providing right authentication. Apart from the Privacy concern, drawing data from the database and Use it in a way to provide the right kind of prescription was a challenge but Artificial intelligence played significant role in achieving the desired result. Measuring Project's success is depend on various parameters like performance, efficiency and data interoperability. Since the application works on the android platform. The app has to check whether the device satisfies all the minimum requirements of memory size, processor speed, accessibility to various Application programmable interfaces (API). The central parameter of success is the accuracy at which the machine learning algorithm predicts the diseases for which it is trained. It is expected to be above 96% -98% to make a real difference, as medical diagnosis has very rigorous standards The secondary parameters are the response and loading time of each modules. A high value of module latency can make any user impatient. Besides these 2, the other parameter is the number of features that work without any bugs. This parameter should be ideally above 90% as well, or it can make users frustrated. Sharing Data between different platforms is also one parameter as the app is supposed to collect data from the user input and generate result based on the past history and results it acquired.**

I. INTRODUCTION

Mobile healthcare applications (Apps) gets more attention due to the ability to modernize healthcare delivery, for example, patients can check their daily activity and habits on Mobile Healthcare apps by using android, it gives them the advantage of solving their query of their symptoms and get an expert response from the Expert System in the form of identification of the disease and medication the cure their illness. Patients gets an access to the medical information and expert system independent of their current place and time and content can be dynamically adjusted to the current context and terminal type. This proposed system, is designed to partially ease the stress, financial burden, and workload on the doctor.

This application which runs on any Android version. This apps and service usability has to be user- friendly, and usefulness of the system due to great spectrum of potential future users and variety of their needs and expectations from system's functionalities. While the session is transferred, an additional mechanism session consistency must be done to protect storage and to avoid transfer of invalid session data. Due to the various network varieties that may be used, it's extremely attainable that communication is interrupted unexpectedly which a system should be ready to save current session information and change a user to continue antecedently started session while not data loss.

When the applying is resumed on a similar terminal over a unique network or on a replacement terminal, a replacement authentication method should defend security. as a result of this could happen oftentimes, it's necessary that the revived authentication is each easy and secure. professional systems are being progressively utilized in medical environments like hospitals, laboratories, and medical aid units, with a read to rising the standard of health care and reducing the chance of incorrect medical selections. Transformation of those systems into mobile solutions would extend their edges and facilitate their integration into medical environments. New generations of mobile devices supply user new modes of interaction with health worker systems.

Our approach involves the utilization of a mobile client-server model using net services so as to transfer the presently offered web-based system onto associate automaton platform. The server is devoted to supply associate interpretative report of the obtained check results, whereas the shopper acts as a convenient user front-end. Communication between the shopper and also the server relies on net services. The patients have their user accounts on the patient web and may access the service from distant places or their homes, cars or offices victimization differing types of networks (e.g. Wi-Fi, GPRS, 3G).

II. PROPOSED WORK

Pervasive healthcare system makes use of wearable medical sensor with wireless interconnections to monitor patient's condition and to provide healthcare services for

patient wellbeing. This dissertation is aimed at investigation and development of secure and pervasive architectural prototype for patient monitoring and remote medical care which reflects the unique and efficient wireless solution for monitoring patient in a secure way with a solution of various technological issues. Most important part of pervasive healthcare system is the Body Sensor Network (BSN) which is a suitable combination of wearable tiny devices attached to the patient's body to monitor patient's physiological data (or BSN data). Sensors continuously monitor and collect patient's data and send it to a remote server through a wired or wireless network. This server can be called Database Server (DBS).

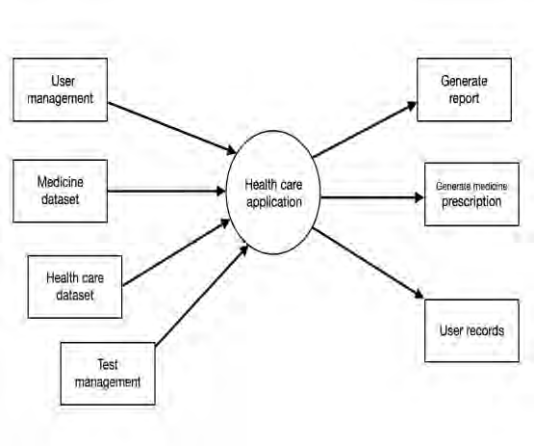


Figure 1

III. LITERATURE SURVEY

1. AI Using AI to improve hospital inpatient care, Daniel Neill, (2013) - Adopt best practices for preventive care and patient safety have huge and directly measurable impacts on both healthcare costs and patient outcomes.

   Solution: Clinical decision support systems (CDSS) were one of the rest successful applications of AI.

   Research Gaps: Practice management, Presence of DSS.

2. AI in healthcare and research, Natalie Banner, (May 2018) - A range of healthcare and research purposes, including detection of disease, management of chronic conditions, delivery of health services, and drug discovery.

   Solution: The app Ada Health Companion uses AI to operate a chat-bot, which combines information about symptoms from the user with other information to offer possible diagnoses.

   Research Gaps: Detection of disease, using chat bot in AI.

IV. FEASIBILITY STUDY

The project relies on open source software, such as python, R, NumPy and SciPy. The functionality provided by these software's is enough to construct a product capable of usage for healthcare applications.
For the machine's learning and training, we will use freely available datasets. These sets run into thousands of lines, and thus can provide enough learning capability to the software. Due to these 2 being freely available, the project is very low cost and thus feasible from an economic standpoint. The expected total cost is thus 0 rupees. As such, we do not need any outside funding. The project will be expected to make minimal sales, due to the presence of a variety of other similar tool and the slow adaption of software's by the medical community. The capabilities required for the project are fairly in the feasible range. Most of the planned feature hinge around processing datasets. A GUI is required, which will be done with the PyGUI framework. The GUI will be kept simple and minimalist. For the hospital searching and insurance selection, we will use readily available data. We also have a feature for users to set personal data, preferences. For the machine learning, we will use Neural networks to train our datasets, both of which are available on open source repositories.
On an operational level, the software will need to run intensive processing tasks while it's in its learning phase. Since the learning phase is a very short process, it is doable without the need for extra hardware. The software will be deployed as a Desktop application and will need users to download it. It will also need internet to operate.

V. RESULTS AND DISCUSSION

1. Expected Results

When the user puts in the symptoms, some algorithms like decision tree, naive bayes algorithm are applied on the data set of the user which is been entered by the user by selecting the pre-defined symptoms in the app and there will be some flow of questions and at the end a report is been generated using these algorithms and medicine is prescribed for the same. The application has proven to gain significant and drastic change in the healthcare industry by enabling common people to get their problems or diseases diagnosed at their place and their availability.

Deployment phase is still under construction, modules are to be implemented yet. The GUI by itself is not very difficult to deploy. We have multiple ways of offering interfaces, such as web interfaces and command line interfaces. It will expand to a desktop GUI in the

future as well. The web interface will be deployed using heroku or AWS, depending on the relative cost

of hosting. Once all the API's are ready the app will work as expected.

Detailed analysis of the project has been done and vast research in context of literature and previously published articles has been carried out. The outcome remains the emerging progress in the field of healthcare using the application. Providing the range of healthcare and research purposes, including detection of diseases, management of chronic condition, delivery of health services and drug recovery along with the best possible prescription available. Documentation of the entire project has been done and all the prerequisites been maintained.

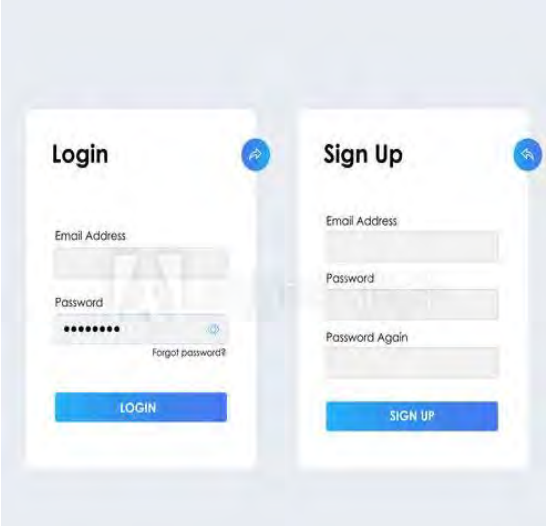


Figure 2

Discussions

Detailed analysis of the project has been done and vast research in context of literature and previously published research papers of IEEE has been carried out. The outcome remains the emerging progress in the field of healthcare application using Artificial Intelligence. Providing the range of healthcare and research purposes, including detection of diseases, management of chronic condition, prescribing medicines.

VI. CONCLUSION AND FUTURE SCOPE

AI technologies have the ability to help address significant health challenges and the strength to overcome every possible limitations and problems where the humans can generally inhale more efforts and time. The application has proven to gain significant and drastic change in the healthcare industry by enabling common people to get their problems or diseases diagnosed at their place and their availability. Henceforth this will result in saving lot of time for both, the patient & doctor and also saving a handy amount which the common people usually invests in getting their tests done on the basis of doctor's Doubts. Because of this innovation, the laboratory tests will become inexpensive and reduction in prices will be observed. This application will have a remarkable impact on healthcare industry.

As per the literature survey we have some deterministic goals for our android app which are as follows; getting the high accuracy rate, flow of algorithms, correct medicine prescription for the diseases, graphic user interface. The developments of the project in the future are aimed at careful consideration of the disease. This can make a prediction of the diseases by looking at a medical history of a number of patients. There after, the team intends to build a web framework to create a rich user interface.

VII. REFERENCES

[1]. jiang F, Jiang Y, Zhi H, *et al* Artificial intelligence in healthcare: past, present and future *Stroke and Vascular Neurology* 2017;2: doi: 10.1136/svn-2017-000101

[2]. Daniel Neill, Using AI to improve hospital inpatient care, 5 oct 2013.

[3]. Natalie Benner, AI in healthcare and research, May 2018.

[4]. Williams K, et al. (2015) Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases J R Soc Interface 12: 20141289.

[5]. Alder Hey Children's NHS Foundation Trust (2016) Alder Hey children's hospital set to become UK's first 'cognitive' hospital.

[6]. Dilsizian SE and Siegel EL (2013) Artificial intelligence in medicine and cardiac imaging Curr Cardiol Rep 16: 441; Future Advocacy (2018) Ethical, social, and political challenges of artificial intelligence in health.

[7]. Written evidence from Royal College of Radiologists (AIC0146) to the House of Lords Select Committee on Artificial Intelligence; Hainc N, et al. (2017) The bright, artificial intelligence-augmented future of neuroimaging reading Front Neurol 8: 489.

[8]. Wang D, et al. (2016) Deep learning for identifying metastatic breast cancer arXiv preprint arXiv:160605718; Esteva A, et al. (2017) Dermatologist-level classification of skin cancer with deep neural networks Nature 542: 115; Rajpurkar P, et al. (2017) CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning arXiv preprint arXiv:171105225; Moorfields Eye Hospital (2018) DeepMind Health Q&A.

[9]. See http://www.ultromics.com/technology/. 30 Bedi G, et al. (2015) Automated analysis of free speech predicts psychosis onset in high-risk youths, NPJ Schitzophrenia, 1: 15030; IBM Research (5 January 2017) IBM 5 in 5: with AI, our words will be a window into our mental health.
[10]. Kassahun Y,et al. (2016) Surgical robotics beyond enhanced dexterity instrumentation Int J Comp Ass Rad 11: 553-68.
[11]. Medical News Bulletin (20 January 2017) Artificial intelligence app Ada: your personal health companion; see also: https://ada.com/.

SMART WARDROBE – IOT BASED APPLICATION

Jay Dalal, Anshul Dalmia, Jaimin Desai
Information Technology, Thakur College of Engineering and Technology, Kandivali.
Mumbai, India

Abstract – **Every individual faces a problem of what to wear, almost every day. The problem isn't that you don't have enough outfits to wear, but fact that you have no track of the clothes you keep wearing again and again and those you haven't touched for a long time. A smart wardrobe not only helps you track your clothes but also get suggestions on what to wear according to the events of your calendar. The system generates statistical reports of the frequency of usage of the clothes which helps you decide the ones you could sell or donate. You can also manage your clothes from anywhere with a mobile application.In this thesis project, our goal is to put together a set of technologies into a system that could be used to aid the user in the planning of their clothes. In order to achieve this goal, we need to be able to identify which clothes are worn very often and those which are seldom used. We need to categorize our clothes like Casual, Formal, Dinner Party, Business Party, Social Function and so on. The application will provide statistics to show the clothes that are being worn frequently and those that haven't been touched for a long time. Application can further suggest the user to sell the unwanted clothes or donate them to charity.Smart Wardrobe is a fairly new area of research and literature on the subject is not readily available or accessible, whatever that exists is mainly conference proceedings or reports on projects most of which have not been progressively developed or didn't get to be marketed commercially. The components include an Raspberry Pi 3 Model B, Arduino UNO, RFID system and Microsoft Azure Cloud assisted by a middleware application working together to give the user a new experience on how they interact with their wardrobe and at the same time enable the wardrobe to support the user in selecting their clothes to be wear in different occasion thereby saving their time they would have spent in the manual Selection and at the same time reducing on the stress associated with the activity.In the future, this project can be extended to provide suggestions to the user according to the weather forecast. Weather detecting sensors can be embedded into the system to detect the weather and suggest appropriate clothes. The app will also give more detailed analysis of the statistical data and machine learning will be used to provide better suggestions.**

I. INTRODUCTION

Today, individuals have access to an array of smart applications like smartphones, laptops, smart televisions, refrigerators, microwave ovens, etc. However, there aren't many smart devices for the bedroom department of the house. One prominent and important feature that exists in the bedroom is a wardrobe or closet. In fact, over 80% of the respondents we surveyed agreed to have access to a wardrobe. Smart wardrobe helps users manage their clothes inside their wardrobe. Besides having an attached screen to show the status of each cloth, it can also push all data into a cloud and further sync up data with a mobile application. User can get suggestions on what to wear today from their phone app based on the events in their calendar and the weather.In this thesis project, our goal is to put together a set of technologies into a system that could be used to aid the user in the planning of their clothes. In order to achieve this goal, we need to be able to identify which clothes are worn very often and those which are seldom used. We need to categorize our clothes like Casual, Formal, Dinner Party, Business Party, Social Function and so on. The application will provide statistics to show the clothes that are being worn frequently and those that haven't been touched for a long time. Application can further suggest the user to sell the unwanted clothes or donate them to charity.

II. MATERIAL AND METHODS

Materials:-

1) Hardware components:

i) Raspberry Pi 2 Model B

ii) Arduino UNO

iii) JT2850 RFID module

iv) JT606 RFID antenna

2) Software apps and online services:

i) Microsoft Windows 10 IoT Core

ii) Microsoft Azure

iii) Service Bus, Web Apps, SQL DB

iv) Arduino IDE

v) Microsoft Visual Studio 2015

vi) Android Studio

3) Hand tools and fabrication machines:

i) Microsoft PC running Windows10 with Visual studio 2015 installed

ii) Android Phone

iii) Raspberry Pi USB Wi-Fi dongle

Methodology:-

Phase I:

1) Planning: Applying agile methodology for Smart Wardrobe for better usage of resources and time with great performance.

2) Analysis: Prepare a detailed analysis on Smart Wardrobe system and overcoming its limitation in our system.

3) Design: Integration of data and designing of system.

4) Coding: Writing the code for RFID module and taking help of open source.

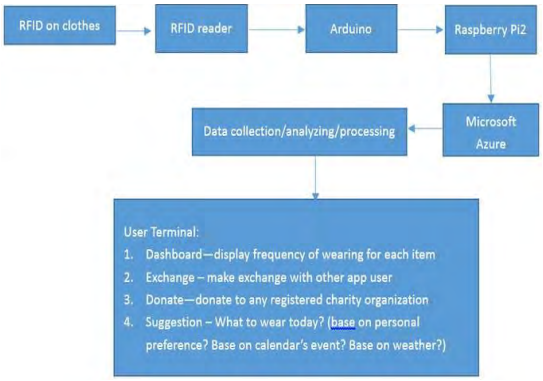
5) Implementation: For first time registration, user will need to let RFID reader read the tag and a push notification will be shown inside user's mobile phone application. User will need to key in further details about his cloths for example color, type, name, etc. Once record is created, system will automatically track the item taking out or putting in from wardrobe. After consolidating data, a report will be generated and give user better idea of which cloth is most preferred or which is most unwanted.

Phase II:

1) Testing: Doing the various tests on smart wardrobe system like unit testing, integration testing, and system testing etc. of test case data to check if the integrated system functions are as desired by the client.

2) Deployment: Conduct beta testing for identifying any further errors, bugs and improvements that can be performed. After the testing and approval, deploy the proposed system.

Workflow:-



The project will consist of 4 main components: user terminal (phone, pc, tablet), Raspberry Pi2, RFID reader, and Microsoft Azure. For first time registration, user will need to let RFID reader read the tag and a push notification will be shown inside user's mobile phone application. User will need to key in further details about this clothes for example colour, type, name, etc.

All clothes stored inside wardrobe has a RFID tag attached to it. Whenever the clothes are put in or taken out from wardrobe, they will be tracked by RFID reader and arduino board. The sensor data will then send over to Raspberry pi 2. After processing the data, Raspberry Pi2 will then send the data to Azure SQL database.

Once wardrobe data is sync up with universal app, the app will further show user which cloth is most frequently chose and which one has the least frequency.

III. RESULT

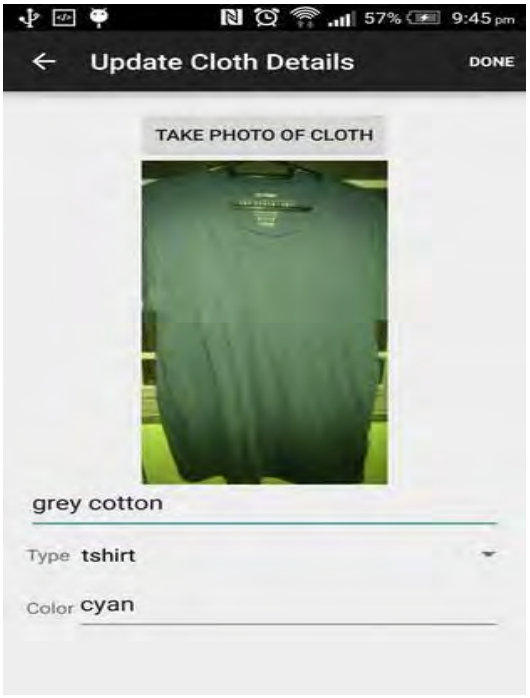
Expected Outcome:-

-Phase I:

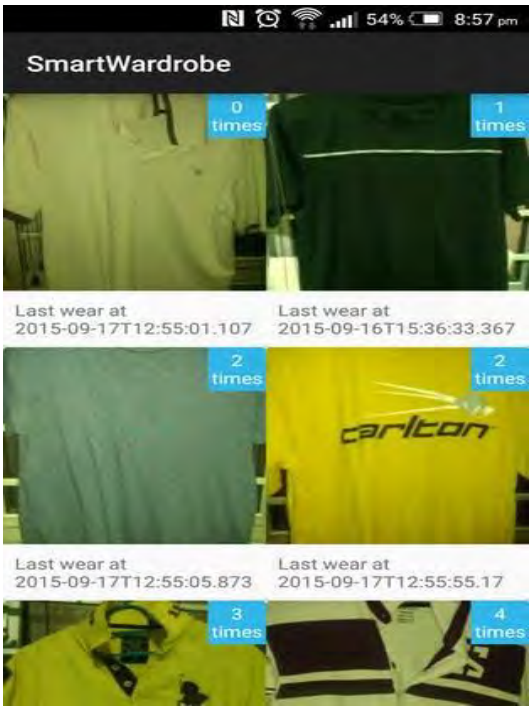
This system will be instrumental into keep tracking of wearing frequency of cloths in wardrobe. It will also suggest a garment on basis of least recent used phenomenon. This system will store usage data and log data on Azure Cloud, this will furnish user to access data from anywhere anytime.

-Phase II:

Any bugs later found will be dealt with and new features like suggestions through weather tracking will be implemented.



To add a new cloth to app



Cloth detailsFrequency of usage

IV. DISCUSSION

In the future, this project can be extended to provide suggestions to the user according to the weather forecast. Weather detecting sensors can be embedded into the system to detect the weather and suggest appropriate clothes. During rain, the system may remind the user to carry an umbrella and avoid wearing white, while on the other hand, during a sunny day, the app suggests the user to wear light colored clothes. The app will also give more detailed analysis of the statistical data and machine learning will be used to provide better suggestions. The hangers could be equipped with LED lights that illuminate at the search of a particular cloth making it easier for the user to locate it in the wardrobe. One of the most important thing aspects to consider in the future is to reduce the effective cost of the system so that everyone can affordit.

V. CONCLUSION

Thus by this we conclude that the proposed system for smart wardrobe can be used effectively by fashion bloggers, organization etc. Not only an organizer but a normal person can use the system as it is simple to understand. As is, the system may require quite some changes to realize a user experience that would be acceptable but at least we have provided a basis for future research. We have identified challenges and future enhancement.

References

I] Paper references:

[1]. [1]Adomavicius, G. and Tuzhilin, A. (2001) 'Using data mining methods to build customer profiles', Computer, Vol. 34, No. 2, pp.74–82.

[2]. [2]Amato, G. and Straccia, U. (1999) 'User profile modeling and applications to digital libraries', Proceedings 3rd European Conf. Research and Advanced Technology for Digital Libraries, Springer-Verlag, pp.184–197.

[3]. [3]Balabanović, M. and Shoham, Y. (1997) 'Fab: content-based, collaborative recommendation', Communications of the ACM, Vol. 40, No. 3, pp.66–72.

[4]. [4]Chen, H.C. and Chen, A. (2001) 'A music recommendation system based on music data grouping and user interests', Proceedings of the 10th International Conference on Information and Knowledge Management, ACM Press, Atlanta, pp.231–238.

II] Web references:

[1]. www.hackster.io/seeyouu/smart-wardrobe

[2]. www.digitaltrends.com/home/threadrobe-smart-wardrobe

Smart Bag Using IoT

Sneha Jaiswal, Akshara Lagwankar, Ms. Payal Borchate

Department of Information Technology, Thakur College of Engineering and Technology, Mumbai, India.

Abstract – **The proposed system uses the technique of Internet of Things in order to track the bags. In this, hardware would be created and installed which would be having the basic arduino board with a GPS module and an alarm being connected to it. A map has been created which would be synchronized in order to track the location of the bag.**

Furthermore, the map has the features that as soon as the bag gets lost or theft and it moves away from the owner and goes out of a particular range, the alarm would start ringing so that the owner gets notified where exactly the bag is. Also it would help the owner to track down the location of the bag which could be seen on the map as the markers would be dropped which gives us the location of the bag as it moves away from the owner.

Keywords: Raspberry Pi, Bluetooth, GPS, Android application

I. INTRODUCTION

When people travel, they travel with their luggage anywhere they go so it is what a part of their journey becomes. Baggage however when carried needs to be carried to places and with great care because it contains many important and necessary things. In today's world as we know the amount of theft and forgery has been considerably increased hence there is a need to keep our dear ones and our belongings safe. However, what would happen if this luggage gets lost or robbed in the middle of your journey it would be a menace because it would contain all your important things or sometimes documents. Hence, to provide a solution to this problem would include introducing the concept of smart bags thereby providing a better security. It is one of the alternative solutions where instead of looking after your baggage every now and then where the passengers follow the bag, the bag follows them.

I. PROPOSED SYSTEM

The proposed system uses the technique of Internet of Things in order to track the bags. In this, hardware would be created and installed which would be having the basic arduino board with a GPS module and an alarm being connected to it. A map has been created which would be synchronized in order to track the location of the bag. Furthermore, the map has the features that as soon as the bag gets lost or theft and it moves away from the owner and goes out of a particular range, the alarm would start ringing so that the owner gets notified where exactly the bag is. Also it would help the owner to track down the location of the bag which could be seen on the map as the markers would be dropped which gives us the location of the bag as it moves away from the owner.

II. METHODOLOGY USED

Agile methodology mainly focuses on the incremental and the iterative development of the different project phases. It controls and determines how the various stages are implemented and what kind of environment is involved. The project is first divided into different phases, and then the responsibility of developing these phases is given in the hands of the various developers. At every stage of the project development it is continuously monitored and taken care of. These developers developing the phases have particular amount of time period and tasks involved with them. To remain in sync with the other members of the team they have to complete their task within the stipulated amount of time.

The meetings however are conducted on a regular basis where all the members of the team have the right to put forth their ideas along with any requirement or concern. The product backlogs are

III. FEASIBILITY STUDY:

a. Executive Summary

Project is related to Smart Bags.

i. The project maintains two levels of users:-

1. Administrator Level-Owner
2. User Level-Data Entry Operator

ii.Main facilities available in this project are:-

1. Maintaining the daily time-table.
2. Providing temperature sensing.

discussed in these meetings where the requirements of the project to be completed along with the resources are mentioned. However, during the development of the project it is bound to have certain changes. These changes are discussed and team is required to document it along with their consent. Some changes involved may be bad for the future but certain changes turn to be good in the overall run of the project. The changes can be from the customer side or the developers' side. Agile is considered to be a methodology that is ready to accept and work on any change that occurs during any phase of the project which can be from any side of the project. It is flexible methodology that welcomes change during any stage of the project. After the development phase each task or each developed phase is tested in different environment. This helps the team members to determine what would be the next phase of the project. Every iteration of the project involves a testing phase. It requires close communication with the project master along with team members and the product owner for the development of the project

3. Providing GPS location tracking.
4. Reminding customers of the luggage.
5. Providing the weight capacity of the bag.
6. User can set a reminder for his/her different requirements.

b. DEFINE BUSINESS PROBLEM OR OPPORTUNITY

The Software is for the automation of Smart Bags. It maintains one level of user:-

i.User Level The Software includes:-

ii.Maintaining customer details.

iii.Providing and maintaining all kinds of storage facilities for a customer.

iv.Reminding and giving an alert to the user as per requirements.

c. REQUIREMENTS AND PURPOSE OF THE STUDY

i.Hardware Requirements

- Raspberry pi, Virtual Assistant
- Bag

ii.Software Requirements

iii.Android Studio

- *Purpose*
- Delightful Bags provide following facilities like:-
- Taking orders from the customer.
- Delivering ordered meal timely
- Bill computation
- Record maintenance.

d. DESCRIPTION OF THE OPTIONS ASSESSED

Based on research conducted over the Internet we identified the following alternatives for the MMS:

- Continue to use the current system.
- Utilize a packaged solution. The bag providers who require this software can purchase the product and host them in their own environment. Less amount of programming and re- programmingisrequiredfor customizing the system and implementing them.
- Develop a customized bag solution which meets all requirements satisfactorily and is also built with a scalable scope.

e. ASSUMPTIONS USED IN THE STUDY

Every user should be:

•Comfortable bag usage.

- He/she must have proper internet connection.
- He/she must also have basic knowledge of English too.

f. AUDIENCE IMPACTED

It can be used in anywhere as far the internet services are provided.

g. FINANCIAL OBLIGATION

The financial and the economic questions during the preliminary investigation are Verified to estimate the following:

i. The cost to conduct a full system investigation.
ii. The cost of hardware and software for the class of application being considered.
iii. The benefits in the form of reduced cost.
iv. The proposed system will initially give the minute information, as a result the performance is improved which in turn may be expected to provide increased profits.

II. HARDWARE DESCRIPTION

The various components required for the development of the project Smart Bag using IOT along with their estimated budget is as shown below:

- RASPBERRY PI 3 MODEL B INBULT BLUETOOTH AND Wi-Fi.: The main component for designing of the system for internet based system.
- Intel i3-2120 3.3GHz Processor: Computer system processor
- HITSAN 3x Heat Sink Kit +Transprent Acrylic Case+Cooling Fan For Raspberry Pi
- SunFounder USB 2.0 Mini Microphone for Raspberry Pi 3, 2 Module B & RPi 1: microphone for interaction between the system and the users
- Model B+/B Laptop Desktop PCs Skype: optional for advance use
- GSM MODEM MODULE BOARD WITH SMA ANTENNA - CALL SMS GPRS facility
- USB GPS receiver module BMP180
- Digital Pressure Sensor Raspberry Pi
- BS-708 UBX-G7020 USB GPS receiver module with antenna replace BU-353S4: For location tracking facility
- BU353S4 - suitable for laptop and automobile
- BMP180 Digital Pressure Sensor Arduino Raspberry Pi Compatible Sensor for Senior
- Design Capstone Design Project

DESIGN AND IMPLEMENTATION

5.1 *Data Flow Diagram (DFD)*

DFD level 0:

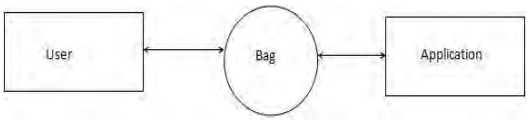


Fig 5.1.1

DFD level 1:

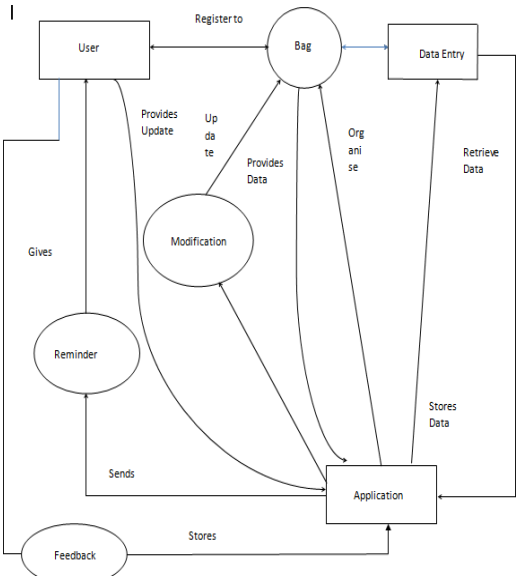


Fig 5.1.2

DFD level 2

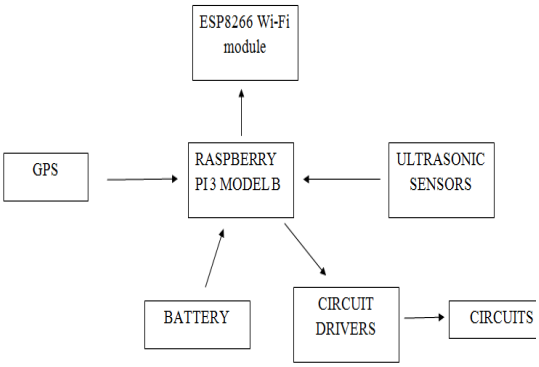


Fig 5.1.3

5.2 Block Diagram and flow chart:

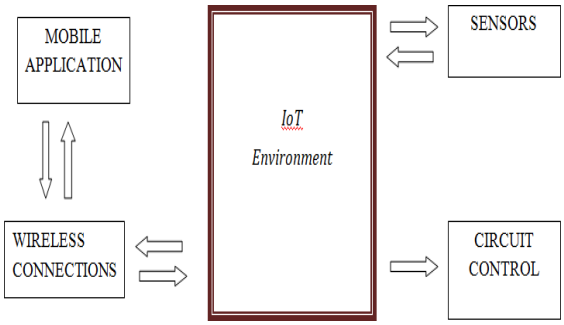


Fig 5.2.2

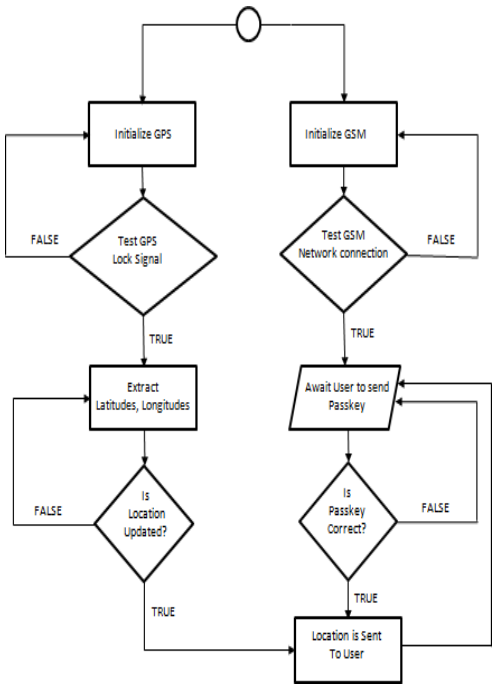


Fig 5.2.1

6.Future Scope

In future, we are planning to include some interesting features like automatic object avoidance, stair case climbing and some extra woman safety features. These extra features make the bag more powerful and user friendly.

IV. CONCLUSION:

The progress in science & technology is a non-stop process. New things and new technology are being invented. As the technology grows day by day, we can imagine about the future in which thing we may occupy every place. The proposed system based on microcontroller is found to be more compact, user friendly and less complex, which can readily be used in order to perform several tedious and repetitive tasks. Though it is designed keeping in mind about the need for school kids, it can extended for other purposes such as commercial & research applications. Due to the GPS tracking working logic: probability of high technology (Android) used this system is well software controlled. The feature makes this system is the base for future systems. In future we can use this idea of smart bag to packing system in factories, shops, super markets etc.

This project idea shows the implementation of tracking the bags which are either lost or stolen using IoT. Certain procedures and techniques have been made and proposed in order to achieve the same. Experimentation has been done and maps have been created in order to track the location of the bags which are misplaced and lost.

Experiment results further concludes that the bags can be easily tracked based upon the hardware installed in them and then by tracking that hardware and tracking the route, directions and location of the bag with the help of a map.

References:

[1]. SMART BAG- 1. Mrs. Rasika Naik, 2. Smt. Sanjana Mupidwar,3. Pallavi Chavan, 4. Pooja Chindarkar, Department of Electronics and Telecommunication, Vivekanand Education Society's Institute of Technology, Mumbai, India 2016.

[2]. Smart Bag Tracking and Alert System using RFID: Shubham Sarkar, Suvojit Manna, Subhadeep Datta, 2017.

[3]. LUGGAGE TRACKING SYSTEM USIN IoT: Sudha enthikumar, Brindha K, Rathi R., Mayank Jain 2017.

Real-Time Multiple Object Detection on Low Constrained Devices using Lightweight Deep Neural Network

Dr.Vinayak Ashok Bharadi, Namdeo Baban
Research Scholar, IT department FAMT, Mumbai University Ratnagiri, India
vinayak.bharadi@famt.ac.in, namdeobadhe1982@gmail.com

Abstract— **During the past few decades, deep neural networks have attracted considerable attention due to their wide applications in filtering, combinatorial optimization, style transfer, pattern recognition, computer Vision and some other areas. object detection is one of the challenging task in computer vision field. There are many popular deep neural networks based object detection methods. The popular ones are Region-based Convolution Neural Networks (RCNN), Single Shot Multi object Detector (SSD) and You Only Look Once (YOLO).All these models require huge computation and large power consumptions which is a big challenge for real-time multiple object detection deployments on resource constrained devices, e.g., mobile phones or embedded gadgets. The combined version of SSDlite and Mobile Nets gives better result for object detection and tracking on resource constrained devices .In this paper, the SSDlite-mobilenet_COCO models are trained on GPU GTX 1660 GeForce GTX 1660 Ti using MS- COCO dataset and tested on using RabberiPi 3 B+. The performance in terms of Frame per second (FPS) of the models on GPU as well as CPU was tested.**

Keywords—Object Detection, Single Shot multi object Detection, MobileNet

I. INTRODUCTION

Object detection (OD) [1] is the most challenging part of any of image processing applications. Given an input image, OD algorithm outputs multiple objects with a class label and its location(usually in the form of bounding box coordinates).It usually helpful to find real-world object instances like car, bike, TV, flowers, and humans in still images or Videos. It can achieve this by learning the special features each object possesses. Most Common applications of object detection are pedestrian detection in road safety system, self-driving cars, emotion detection from

facial expression, Human activity recognition, object tracking in surveillance System, people counting in crowd areas, relief and rescue operations medical imaging, understanding aerial images etc. The images can have object instances from same classes, different classes or no instances at all.

There are many issues present like the variation in size, dark lighting, occlusion, rotation, low- resolution, appearance, complex background other factors all will seriously affect the object detection

performance. The traditional object detection method such as background subtraction, temporal differencing, optical flow, Kal man filtering, support vector machine, and contour matching [2] are normally used to find the features from images manually. But all are not able to detect object accurately.

During the past few decades, deep neural networks have attracted considerable attention due to their wide applications in filtering, combinatorial optimization, style transfer, pattern recognition, computer Vision and some other areas. Deep neural networks have been proven to be effective in various areas. Deep neural networks have achieved state-of- the-art performance in these fields compared with traditional methods based on manually designed visual features.

Convolutional Neural Network (CNN), a successful Artificial Neural Network-based deep learning algorithm, is used in recognition of visual patterns directly from pixel images with variability [3]. CNNs are like traditional neural network, but with deeper layers. CNNs model temporal and spatial associations of features commonly applied to image classification and face recognition. It uses a set of layered functions (i.e., convolution layers) on every possible sections of an inputs array of values in computing the output.In 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [4], Hinton and his student Krizhevsky [5] had applied CNN to image classification and achieved top5 error

15.3% vs 26.2% of the conventional method.

II. RELATED WORK

Object Detection using Region-based Convolutional Neural Networks(R-CNN) was proposed by Ross Girshick[3]. R-CNN is a special type of CNN that is able to locate and detect objects in images: the output is generally a set of bounding boxes that closely match each of the detected objects, as well as a class output for each detected object. The image below shows what a typical R-CNN outputs:

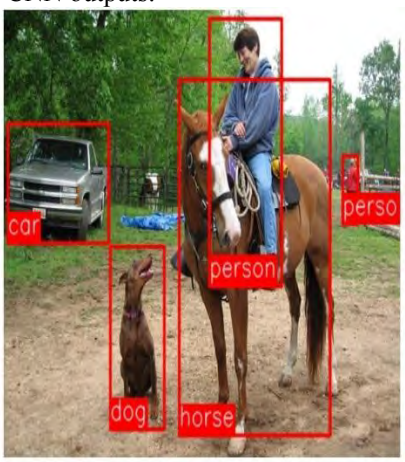


Fig 1 Multiple Object Detection

Region based CNN Models find a bunch of boxes in the image and checks if any of these boxes contain any object. RCNN uses selective search to extract these boxes from an image (these boxes are called regions).There are number of challenges that have to be addressed in order to give a more reliable detection system. R-CNN has problem with respect to heavy and long dataset. Object detection happens in two steps region of interest identification and then actual classification and the performance decreases for unlabeled dataset. R-CNNs does not look at the complete image. Instead, parts of the image which have high probabilities of containing the object.

Regression-based YOLO (You Only Look Once) model was proposed by Joseph Redmon, Santosh Divvala[4].YOLO has a single convolutional network predicts the bounding boxes and the class probabilities for these boxes directly from full images in one evaluation. The problem with YOLO Model is that it only predicts 1 type of class in one grid hence, the performance degrades with small objects within the image, for example detecting a flock of birds.

Single Shot Multi Box Detector (SSD) was proposed by Wei Liu, Dragomir Anguelov[5]. SSD is a multi-scale sliding window detector which slides a local window across the image and identifies at each location whether the window contains any object of interests or not. SSD Model is a modification of the VGG16 architecture. It is

based on a feed-forward convolutional network that produces a fixed-size collection of bounding boxes and scores for the presence of object class instances in those boxes, followed by a non-maximum suppression step to produce the final detections. SSD Model take one single shot to detect multiple objects within the image, while R-CNNS need two shots, one for generating region proposals, one for detecting the object of each proposal. Thus, SSD is much faster compared with two-shot R-CNN Models.

But there are also some limitations of SSD Models like it not only confuses objects with similar categories (e.g. animals) but also produces worse performance on smaller objects, as they may not appear across all feature maps.

Objet detection using Region based CNN, YOLO and SSD models require huge computation and large power consumptions which is a big challenge for real-time multiple object detection deployments on resource constrained devices, e.g., mobile phones or embedded gadgets. A resource constrained scenario means a computing task must be accomplished with limited resource supply, such as computing time, storage space, battery power, etc. Running deep learning models is computationally expensive. SSDlite is a light weight model is primarily used to locate object and Mobilenet classify the object.

SSDlite requires only 2.1 M parameters as compared to Base SSD model (14.8M)[6] SSD_Mobilenet_v2 is a combination of MobileNetV2 and SSDLite can produce the multiple

object detection. MobileNetV2 + SSDLite is 20× more efficient and 10× smaller while still outperforms YOLOv2 on COCO dataset [6]II. PROPOSED WORK

In the proposed system, SSDlite_Mobilenet_v2_COCO model is trained on GPU GTX 1650 Ti server USING MS COCO datasets. It includes all 164K images from COCOs

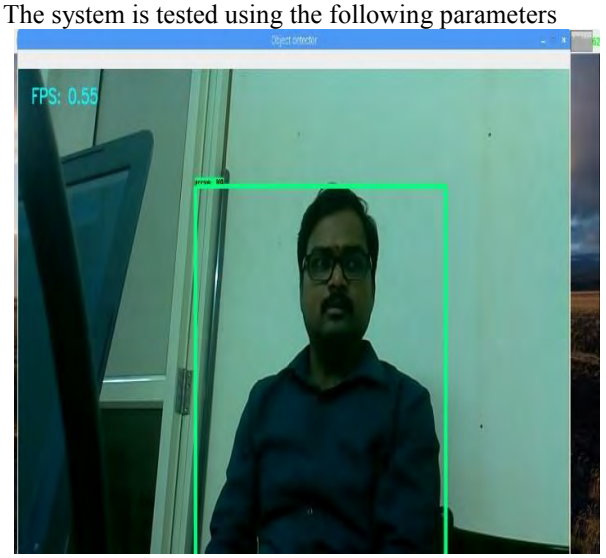
2017 (training 118K, validation 5K, test-dev 20K, test-challenge 20K). It covers 172 classes: 80 thing

classes, 91 stuff classes and 1 class 'unlabeled'.

The trained model is retrained and deployed on RasberiPi 3 Model B has features a quad-core 64-bit ARM Cortex A53 clocked at 1.2 GHz. Input image is taken by a camera is attached to RasberiPi 3 Kit.A model is retrained and results are checked with GPU results.

III. RESULTS & DISCUSSION

The system is tested using the following parameters

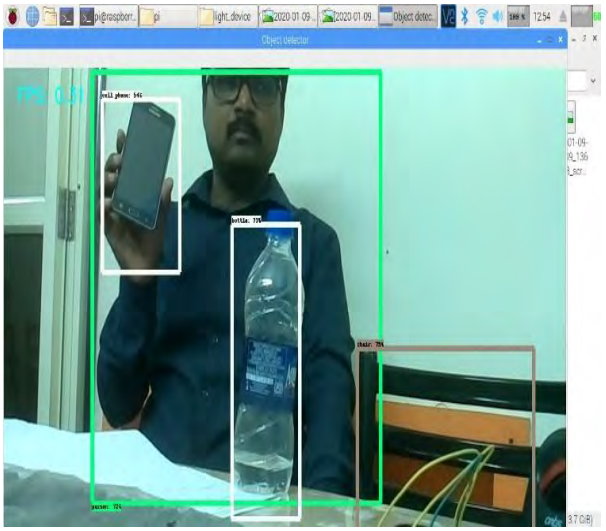


Frame per second (FPS)

Mean average Precision (mAP) The results are pending

Single Object Detection Result

.



IV. CONCLUSION AND FUTURE WORK

The SSDlite_Mobilenet_v2_COCO model is trained on MS COCO datasets. It is deployed on Raspberry Pi 3 Model B. It detects single as well multiple objects in image. The performance of the model on GPU is faster as compared to low constrained devices in terms of FPS. In future, Due to low processing speed and storage space if we run the model for long time, the model restarts again. In future, we will focus on how to improve the FPS by reducing the number of layers and normalization techniques.

References

Multiple Object Detection Result

[1] H. S. Parekh, D. G. Thakore and U. K. Jaliya, "A Survey on Object Detection and Tracking Methods," International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, no. 2, pp. 2970-2978, February 2014.

[2] H. S. Parekh, D. G. Thakore and U. K. Jaliya, "A Survey on Object Detection and Tracking Methods," International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, no. 2, pp. 2970-2978, February 2014.

[3] Ross Girshick, Jeff Donahue,"Region-based Convolutional Networks for Accurate Object Detection and Segmentation"2015 IEEE Transactions on Pattern Analysis and Machine IntelligenceDOI: 10.1109/TPAMI.2015.2437384

[4] Joseph Redmon, Santosh Divvala "You Only Look Once:Unified, Real- Time Object Detection", 2016 IEEE Conference on Computer Vision and Pattern Recognition

[5] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. 2015 International Conference onEngineering and Technology (ICET). doi:10.1109/icengtechnol.2017.8308186

[6] Mark Sandler Andrew Howard," MobileNetV2: Inverted Residuals

Android Employee Tracker

Pranjali Kasture, Aishwarya Nambiar, Shubham Mundhra
IT, Thakur College of Engineering and Technology
Mumbai, India

Abstract—**In the past few years the development and growth of Android based applications have had a great impact on people's lives. The objective of an Android Employee Tracker is to track the location of an employee and track the productivity of an employee. This system or application will automate the tracking and monitoring process of an employee and increase the productivity and efficiency of all the employees. This in turn will increase the overall efficiency and productivity of an organization. The system is a coalition of a web-based application and an android application wherein the employees will use the android application and the management uses the web-based application. This system is intended to monitor field employees where the tracking of an application is necessary in order to monitor the productivity. When the employee opens the android application, the application will take the employee to the login page. The login page will contain the classic login fields such as login id and password. The employee will click a picture with the photo clicking option and submit along with user id and password. This will take the employee to the page where the location in terms of latitude and longitude coordinates are displayed and the field for describing the purpose of the employee's visit.**
The management will receive a notification or an update through the web-based application where they can check the location details of all the places visited by the employee as well as the log details with the time and the picture as proof.
Using this system, the employees will be monitored and hence
improves the productivity of the employees. This type of application and system could dramatically improve the productivity and efficiency of an organization.

Keywords—location, GPS, photograph, log details, maps.

I. INTRODUCTION

Based on the previous experiences such as inconsistency in data, loss of data, problems in tracking employees, problems in scheduling future activities for the Employees, we are implementing a new generation Android Employee Tracking System. The aim of the Android Employee Tracker application is to automate the employee tracking and monitoring process and improve the productivity and efficiency of the employees and management. This is an essential step in improving the organizational growth of a company and helps improve its overall productivity.

II. PROBLEM DEFINITION

The rapid growth of android based applications are creating a great impact on our lives. The aim of the Android Employee Tracker application is to automate the employee monitoring process. This approach also helps in improving the organizational growth of a company and also increases the overall productivity of the organization. This system helps the administrator and Human Resources department to keep track of the employees who go for field work. Since GPS location of the employee is tracked, so employee will not attempt to add proxy attendance. The system should be robust and be able to make productivity report. The system should be able to take input from the employee's application for tracking purpose. The system should be able to maintain the records of all employees.

Phase wise planning of the system

Planning: Applying agile methodology for tracking employees for better usage of resources and time with great performance.

Analysis: Prepare a detailed analysis on present tracking
systems and overcoming its limitations in our system. **Design:** Integration of data and designing of android and web application.

Coding: Writing the whole system app code and taking help
of open source.

Implementation: Giving user to test alpha testing and gathering centric analysis of performance, feedback, and try to improve the quality of result/output.

Testing: Doing the various test on the system app like unit testing, smoke testing, stress testing, integration testing, regression testing, system testing, etc. of test case data to check if the integrated system functions are desired by the organization.

Deployment: Conduct beta testing for identifying any further errors, bugs and improvements that can be performed. After testing and approval, deploy the proposed system.

III. BACKGROUND

Android Employee Tracker is an advanced monitoring technique which uses android application and web application for automating

the entire monitoring process in an organization or company [1]. This system increases accuracy in managing employees, manager and company's data. This application is important in the global environment to find out the strategic trend of resource utilization in the organization [2]. This system is a combination of a web application and android application where the employee (user) will be usin the android application and the administrator as well as the Human Resources department will use the web application [3]. This system is meant for employees who work in the field and need their location to be tracked. The main reason behind this application is Managing and keeping track of employees is cumbersome and problematic [4]. The Android Employee Tracker assists in the management, monitoring, searching, tracking, updating of employee's record.
where admin will view image and GPS location in web application.
After Login, GPS location of the employee will be tracked automatically by the system and send to the admin after every 5 minutes.
When employee logs out of the system again the image will be
captured as well as GPS location will be send to the admin.

IV. METHDOLOGY USED

Agile methodology mainly focuses on the incremental and the iterative development of the different project phases. It controls and determines how the various stages are implemented and what kind of environment is involved. The project is first divided into different phases, and then the responsibility of developing these phases is given in the hands of the various developers. At every stage of the project development it is continuously monitored and taken care of These developers developing the phases have particular amount of time period and tasks involved with them. The agile methodology used in the development in Android Employee Tracker is based on regular communication with the client. The agile methodology takes into account flexible planning, customer involvement, constant evaluation and risk management. The website development depends on the requirements of the client and the quality of communication with them. The project is developed using Android Development Toolkit(ADT) along with Visual studio 2010 and SQL.

Feasibility Study:

1. Executive Summary

Project is related to Android Employee Tracker
It is a combination of following two applications:
Web application
Android Application

Main facilities available in this project are: -
Maintaining records of employees.
Efficiently tracking employee's location.
Providing field locations to the employees.
Maintaining backup data of all employees.
Updating Employees about the upcoming projects.

2. Define business problem or opportunity

Combination of web as well as android application: User (employee) used android and admin (HR) uses the web application.
When the user will login to the system his image will be captured and his GPS location will be send to the admin

Every user should be:
Comfortable of working with computer
He must have knowledge in developing Android

Application
He must also have basic knowledge of web and development skills

3. Purpose of the study

Purpose

The system produced will be very easy to use
It provides facilities like:

It will be easy to maintain.
The system will be very cost friendly
The system will have interactive user interface

In case any part stops working, all the parts are readily available in the market and are very cheap
The system is based on basic modules which are easy to replace/swap if needed. This shall make it easy to
maintain and to work on.

4. Audience impacted

It is designed for industry with field employees that impact the overall productivity and revenue of the company.

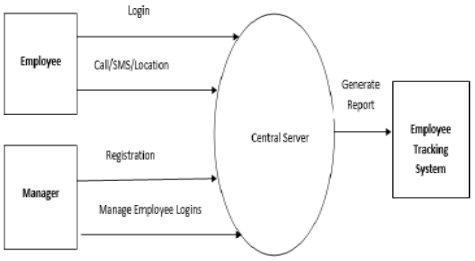
Technical feasibility

As already many employee tracking applications are available but
when it comes to security this is more secure than any other application.

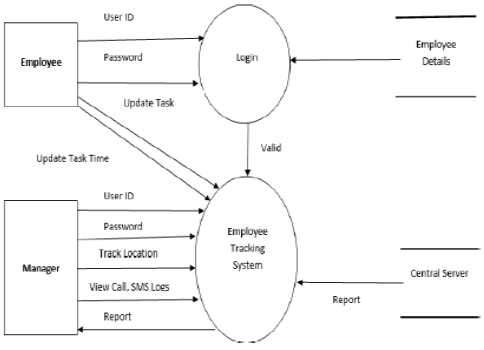
Economic feasibility
As this application is organization specific it will only cost for the
database server to keep backup and host the system on the server.

Socio-cultural feasibility
As Android Employee Tracker is all about managing employees there will be complete transparency between the organization and the employees.V. DESIGN AND IMPLEMENTATION

Dataflow Diagram



F Level 1



Level 2

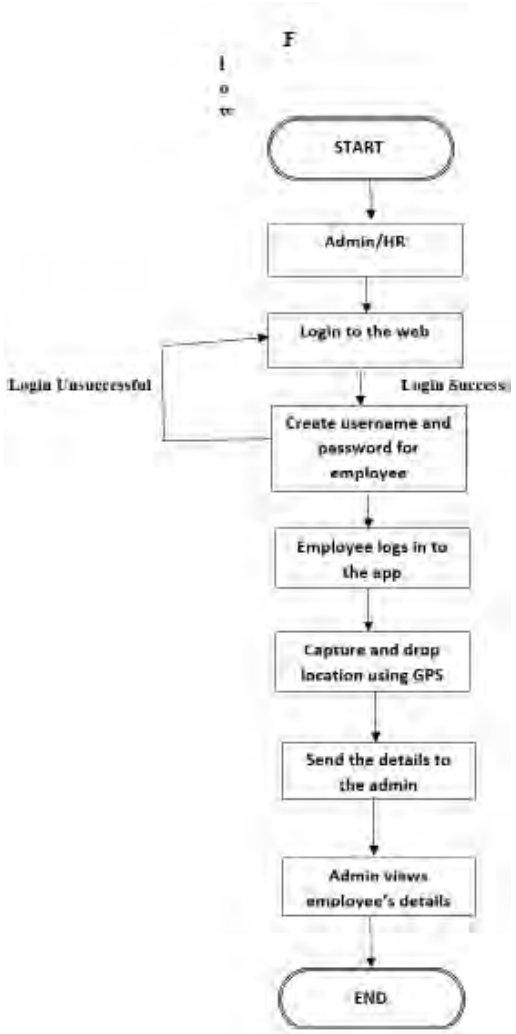


Fig 1: Flow diagram of a system

VI. RESULT

We successfully achieved to eliminate the disadvantages of other messaging systems such as advertisements, distraction
because of friends or other spam messages, fake rumors.
We managed to eliminate the subscription system and instead
of paying in repetitive manner client can own this with one time pay and can get customized according to its needs which is not provided in others.
The latitude and longitude detected by the system has about
90% accuracy wherein the exact location along with neighboring landmarks is visible on the map.

Screenshots:
Web Application:

Fig 2: Login page of web application

Fig 3: Manage Employees

Fig 4: View Location on Maps

Fig 7: View Photograph as proof

Fig 5: Salary Generation

Android Application:

Fig 6: Login Page of Android Application

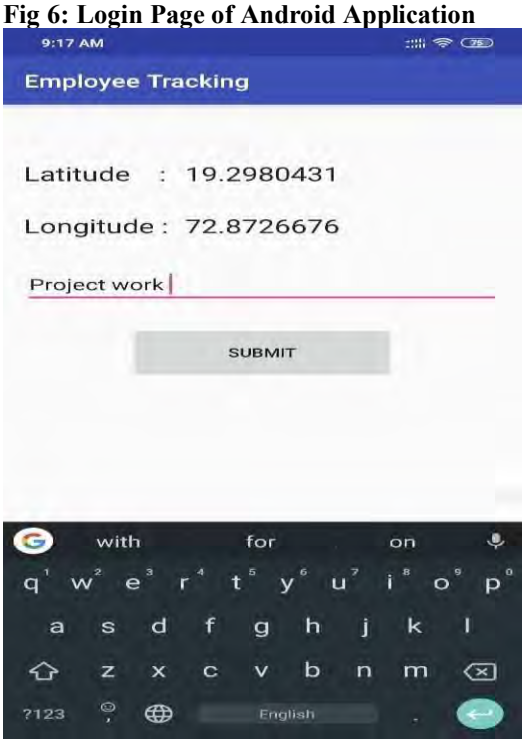


Fig 7: Employee submits location and reason for visit

VII. CONCLUSION:

It can be concluded that with the designed system (Android Employee Tracker), custom tracking can be effectively used by any organization.

Android Employee Tracker application leads to understand tracking using GPS.

Using the web application, the administrator is able to track the employee's location as well as view the subsequent photo uploaded. This web application will also be able to

add, delete or edit an employee's details as well as generate

an employee's salary according to the number of days the

employee was logged into the system.

The system apart from tracking locations can be used to measure the number of working hours and that resulting in the accurate calculation of payroll. A centralized system where all the employees are using one system for in house or out house project, achieving the highest level of efficiency.

This improves the efficiency and productivity of the work done and improves the organization's overall value.

VIII. FUTURE SCOPE:

The system apart from tracking locations can be used to measure the number of working hours and that resulting in the accurate calculation of payroll. A centralized system where all the employees are using one system for in house or out house project, achieving the highest level of efficiency.

References:

[1] IJRAE, 2016 Volume 3, Issue 03,March 2016, "Employee Tracking and Monitoring System Using Android", Sonal Kasliwal, Sushma Kotkar,H.D. Gadade.

[2] IJRITCC,2015 Volume 3, Issue 02, February 2015, "Employee Monitoring System Using Android Smart Phone", by Kalyani Bhagwat, Priyanka Salunke, Shamal Bangar.

[3] IRJET, 2017 Volume 04, Issue 03, March 2017, "Employee Monitoring and Management System Using GPS and Android", by Prashant Dhotre, Ganesh Rasal, Mamta Jha, Nihal Pawar.

[4] IJRCA, 2016 Volume 153, Issue 03, November 2016, "An Android based Employee Tracking System", by Etuk Enefiok A., Onwuachu Uzochukwu C.

Autonomous chatbot building for enhanced user interaction

Vishal Vishwakarma, Vidit Vora, Shubham Singh, Rahul Neve
(Information Technology, Thakur College of Engineering & Technology, India)

Abstract – **Chatbots are being widely used these days to improve the customer engagements and user interaction when the user visits the websites. The applications of chatbots range from HR services, customer support to helping people with bank transactions. The main problem arises when a website owner needs to create his own chatbot based on the website's content. The task of creating a chatbot is very tedious and requires the user to define the intents and entities [4] by himself. Though there are API available in the market, the website owner needs some knowledge about how chatbots work and then he will be able to deploy it. This process requires a lot of time and also data collection from the website is a cumbersome task. The task of gathering data for the chatbot and then feeding it to the API can be automated. We can use a scraper that will gather URLs and text from the given domain name by the user and then feed that data to an NLP engine which will learn on the given data and a basic chatbot which the user can converse with and also deploy will be available in no time. We need to scrap and summarize text and then provide that to the NLP engine. We have explored different NLP algorithms that can be used by the engine on the provided text. This helps us in getting a clear understanding of how NLP algorithms will work and which ones to be used. Majority of the NLP algorithms use POS i.e. Parts of Sentence tagging which is used to key the sentences. A crawler is used to get links from the provided domain name. The crawler is able to gather around 40-50 links based on the domain name.**

The NLP engine is trained on English language beforehand and Analysis of context [1] view needs to be done for the purpose of showing the user what the bot has interpreted from the provided query. This results in reduced number of typed messages and the user is also able to get answer for his query as quickly as possible as the user is able to edit in the dialog flow what the bot thinks and then accordingly proceed with the conversation.

Keywords: Chatbot, automated, bot builder, crawling, entities, intent

I. INTRODUCTION

We aim to create an autonomous chatbot based on the given URL which is provided by the user. The system should be able to get relevant information from the URL/free text that is provided and also form business logic accordingly and that should act as a data for the chatbot. The crawling of web-pages for a given website should be done quickly and then a virtual agent should be ready to answer questions for that given bot. The manual work which is required to build conversational chatbot needs to eliminated to a great extent. The time should be reduced as the user should be able to deploy chatbots on the fly. The cost of this process should be bare minimum as small businesses should be able to deploy affordable chatbots on their websites. Further, to increase customer engagement an attractive dialog flow UI needs to be developed which will have context[1] view i.e. it will show what the bot inferes from a given query of a user and the user should be able to change the value if wrongly interpreted thus helping the user to get his information quickly.

II. MATERIALS & METHODS

Software specifications:- User requirements:
Browser: last stable version of Chrome/Firefox
Developing environment: Operating System: Ubuntu 18.04
Language: Python 3.5, Node js, HTML/CSS for designing the landing page
Frameworks: PyTorch/Tensorflow
Softwares: git (version control), sublime text, pycharm

Hardware specifications (developing environment):-
CPU: 3.2GHz
RAM: >= 8GB
Hard Drive: >=100GB
Network: Wifi/Broadband connectivity

Customer Interaction Details

Website URL: The customer needs to enter his website URL of which he intends to create a chatbot. The URL is compulsorily needed as the system requires the URL to successfully crawl and scrape the website provided for information.

Email-id: The user requires to enter his own email id and then verify it, we need this as malicious user may enter the website name and make the server busy thus causing lag for other users.

Facebook page access token: If the user wishes to deploy the chatbot over his facebook page, we need to

have the access token generated by Facebook for it

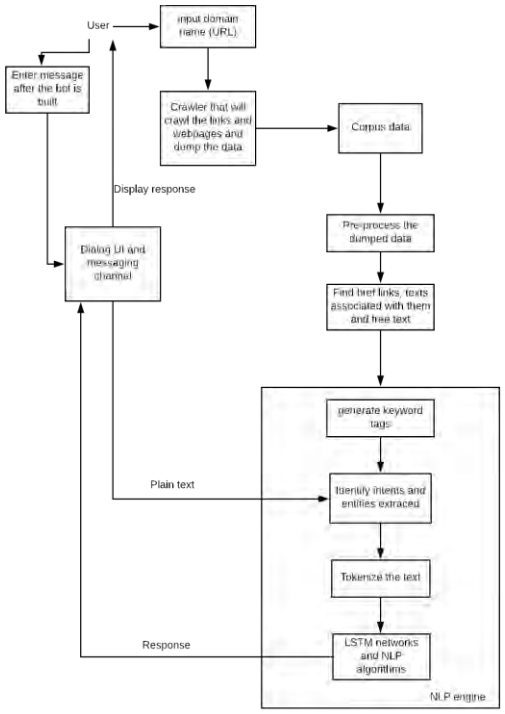


Fig 2.1 : Block diagram explaining the working of the system

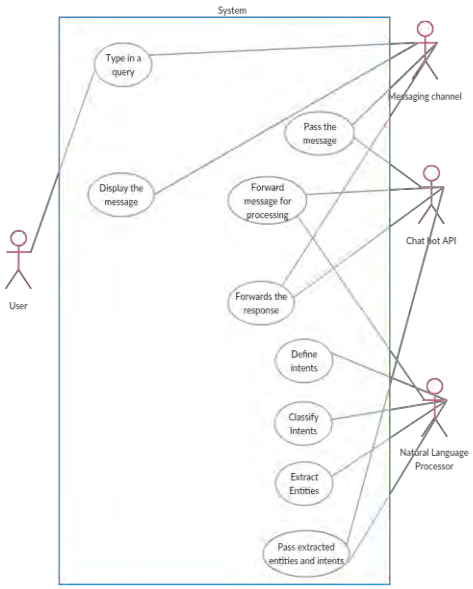
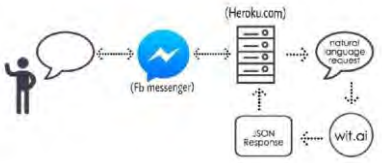


Fig d 2.2: Use case diagram when the system will be eployed.

of of concept to show existing way of creating bots: A sample bot was deployed on facebook essenger after obtaining the links crawled on the llege website. The process of feeding the URL to the was manual. The purpose of doing this was to check ow difficult it is to deploy a chatbot on the facebook en doing it manually.



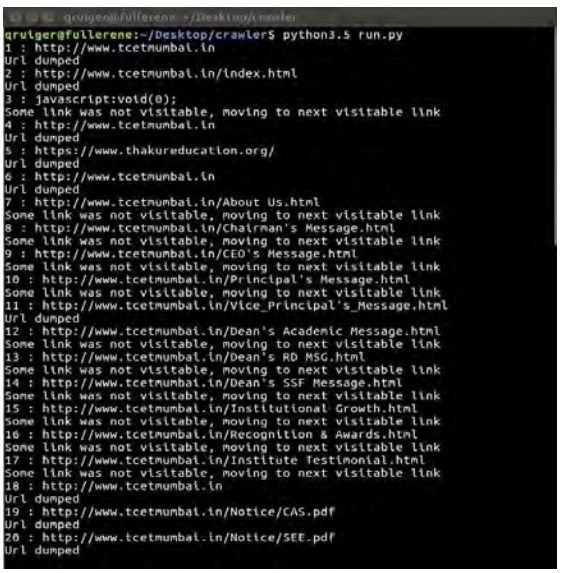
wh

2.3 Flow diagram for the implemented PoC

his process was manual and very cumbersome. The nly automated thing in this process was the crawling of e website. In further phases we plan to make the entire rocess automated as proposed earlier.

III. RESULT

We were able to create a crawler that retrieved links provided the URL of the webpage using existing libraries. The task of surveying various design techniques for conversational agents was also undertaken and the algorithms were analyzed and finalized for our use case. LSTM[6] networks were explored using the Tensorflow API and training the bot first for the English database.



```
qruiger@fullerene:~/Desktop/crawler$ python3.5 run.py
1 : http://www.tcetmumbai.in
Url dumped
2 : http://www.tcetmumbai.in/index.html
Url dumped
3 : javascript:void(0);
Some link was not visitable, moving to next visitable link
4 : http://www.tcetmumbai.in
Url dumped
5 : https://www.thakureducation.org/
Url dumped
6 : http://www.tcetmumbai.in
Url dumped
7 : http://www.tcetmumbai.in/About Us.html
Some link was not visitable, moving to next visitable link
8 : http://www.tcetmumbai.in/Chairman's Message.html
Some link was not visitable, moving to next visitable link
9 : http://www.tcetmumbai.in/CEO's Message.html
Some link was not visitable, moving to next visitable link
10 : http://www.tcetmumbai.in/Principal's Message.html
Some link was not visitable, moving to next visitable link
11 : http://www.tcetmumbai.in/Vice_Principal's_Message.html
Url dumped
12 : http://www.tcetmumbai.in/Dean's Academic Message.html
Some link was not visitable, moving to next visitable link
13 : http://www.tcetmumbai.in/Dean's RD MSG.html
Some link was not visitable, moving to next visitable link
14 : http://www.tcetmumbai.in/Dean's SSF Message.html
Some link was not visitable, moving to next visitable link
15 : http://www.tcetmumbai.in/Institutional Growth.html
Some link was not visitable, moving to next visitable link
16 : http://www.tcetmumbai.in/Recognition & Awards.html
Some link was not visitable, moving to next visitable link
17 : http://www.tcetmumbai.in/Institute Testimonial.html
Some link was not visitable, moving to next visitable link
18 : http://www.tcetmumbai.in
Url dumped
19 : http://www.tcetmumbai.in/Notice/CAS.pdf
Url dumped
20 : http://www.tcetmumbai.in/Notice/SEE.pdf
Url dumped
```

Fig 3.1: Output of the crawler

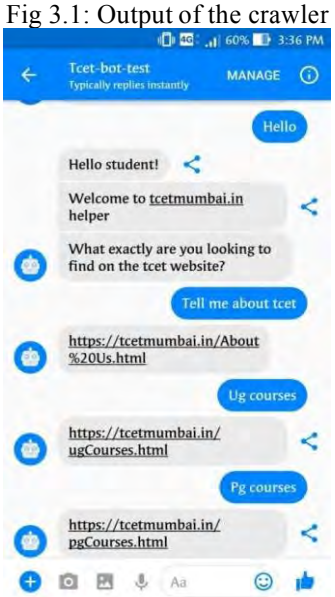


Fig 3.2: Deployed test bot on facebook messenger that replies with crawled URLs

IV. DISCUSSION

The project needs to be diversified by increasing the limit on the number of webpages that is supposed to be crawled. The user should get an option to deploy the bot on a cloud of his choice and a platform that he desires rather than forcing the user to deploy the bot using our cloud server of their choice. The further stages involve working on the GUI of the chatbot and working towards context[1] view, which will hopefully prove to be beneficial to the users that will visit the website owner's website and use the chatbot that was created by our system.

Spelling check in the user query when he enters a message in the chatbot can be done to improve the precision values when the response is generated. The user can have an option to use a dedicated crawler for his own website that will track the changes that are taking place on his website and then update the training data for the bot as well. In further stages, user can just link his FAQ and help pages in the input url and the system will be focused on that only. Thus making a dedicated helper chatbot for the website.

In the current scope there is no scope for user to access how the bot trains itself and the website owner is also not able to create his own personalized conversations that were not able to capture by the crawler used. This will help in filling the gaps that an autonomous bot faces, the learning is not always correct, but the user can intervene to make it correct and then make the committed value as part of the correct values. This will help when the bot encounters the same question again by the user, this time the bot will answer the query correctly. This makes the system a truly learning system.

V. CONCLUSION

Using scrum[8] methodology, we have successfully created a proof of concept wherein the user was able to crawl[7] the links of the provided website and then manually fed to the API. The task of feeding the gathered URLs is not yet automated. A basic command line bot trained to answer basic English questions was also created. We have also explored various NLP algorithms that will be used in the further stages to create the NLP engine. Using Long Short Term Memory networks for chatbots is one of the quickest way to train the bot and it also provides very accurate results. Context[1] view seems to be a very promising feature as during the study it was found that majority of the available chatbots in the market fail to interpret what the user is trying to say but using context[1] view, we will bridge this gap by showing after each message what the bot is interpreting from the user's message and thus enabling the user to edit the view and then proceed with the conversation. This feature is not yet implemented in the mainstream chatbots and thus incorporating it in our project will make it a novelty and also help us to gain insight on how users adapt to these new context[1] view feature and thus helping us to create a comparative study for the same.

References:

[1] Mohit Jain, RamachandraKota, PratyushKumar, Shwetak Patel "Convey: Exploring the Use of a Context View for Chatbots"CHI 2018, April 21–26, 2018

[2] KiranRamesh, SuryaRavishankaran, Abhishek Joshi, K. Chandrasekaran "A Survey of Design Techniquesfor Conversational Agents"Second International Conference, ICICCT 2017New Delhi, India, May 13, 2017 pg. 336

[3] Mao Xuetao, Jean-Paul Sansonnet, François Bouchet "A corpus-based NLP-chain for a web-based Assisting Conversational Agent" May 31, 2014

[4] Vinci Liu, James R. Curran "Web Text Corpus for Natural Language Processing" University of Sydney NSW 2006, Australia [4] https://wit.ai/docs

[5] https://ibm.com/watsom [6]https://www.tensorflow.org/tutorials/sequences/recurrent [7] https://scrapy.org/

[8] https://www.scrum.org/resources/what-is-scrum

Drowsiness Face Detection

Vishal Mishra, Manav Mishra, Kanishk Mishra
Information Technology, Thakur College of Engineering and Technology, Kandivali Mumbai, India

Abstract: **Road Accidents is one of the serious problems seem emerging in the world today. About 1, 05,000 ever year die due to road accidents. It is mainly either due to poor infrastructure or poor driving habits. With the emergence of IT and It's multiple avenues that its offers has proved to be boon to humankind. Drivers Drowsiness is one of the major problems seem emerging and sincere efforts is expected to tackle the problem in order to safe-guard safety of the drive and people in general and also smooth transportation. Detecting real-time drowsiness is done with the help of efficient techniques with consideration of eye and mouth ad primary constraints. It mainly discusses the level of drowsiness and major implementation in consideration. The application is being in C#, python, with OpenCV which uses the webcam which is positioned straight to the face and eye of the driver so as to get the the details of the constraints through eye and mouth. This will enable the extraction of the facial images which will be used to compare with different templates feed through training of the data. It also uses HAAR classifier to extract EAR ratio. The project t is of three types based on physiological, behavioral and the one on vehicular based which is integrated with webcam which is used to alert him upon finding drowsiness while driving.**

I. INTRODUCTION

The project aims at building a face detection system to see the intensity of drowsiness in drivers though system which can be embedded into vehicle system for detection of drowsiness while driving. In the past the number of road accidents have increased by road-side vehicle drivers. The Primary reasons were found to be mainly because of inability to see and inattentiveness. With past studies and researches, it has been of major concern to develop a system which could help drivers while driving. A lot of papers have been published with real time model having drowsiness face detection embedded in them. Amongst them are BMW with Driving Assistant plus, Ford with Active city stop etc.

According to the research conducted this system of vehicle safety helps to reduce and prevent road-side accidents by alerting drivers from getting drowsiness. This project effectively involves studying the drivers' patterns viz. his physical and behavioral patterns which effectively get driver from becoming drowsy. Some of the existing system use steering pattern monitoring, vehicle position in lane monitoring, driver eye/face monitoring etc. The systems available Audi with rest recommendation system, BMW with active driving assistant with attention which analyses the driving pattern which employ graphic symbols. Bosch which takes steering sensor, front mounted lane assists camera vehicle speed turning signal stalk and may other systems available in the list. Ford using driver alert, Honda available with Driver attention monitor Hyundai with its latest Driver Alert system i30.

The model is expected to be implemented in OpenCV python and open-source but as an alternative and ease of language and availability of resource other languages are also used as an alternative. It rings an alarm when drivers found sleeping. This is determining with the help of Famous Jones Viola algorithm also known as (HAAR). Through this the images of eye on its location from face is extracted and with the ratio of blinking proportional to time in terms of delay. The overall aim of the project is to improve efficiency in building the system and reducing the number of road-accidents

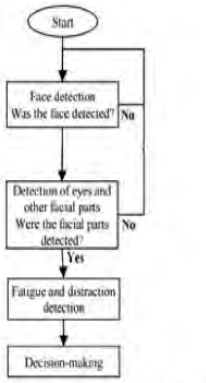
II. MATERIAL AND METHODS

This project application comprises the use of the following resources be it in hardware and software which is discussed as follows:

Hardware specifications:
Sensors
Bluetooth
Webcam

Software Specifications:
Microsoft Visual Studio
Any Text Editor (Visual Studio Code)
Procedure methodology

The output of the Viola Jones Algorithm is a trained Haar Classifier that is stored in a XML File which



Flowchart of the proposed system

can be loaded using the OpenCV function: cvLoad() and then cv Haar Detect Objects() is used to find objects similar to the ones it was trained on to train our own classifiers to detect other objects such as faces , eyes and mouth. This has been done with the OpenCV *haartraining* application, which creates a classifier from a training set of positive and negative samples. Steps are described discussed above as seen in diagram.

III. DISCUSSION

The camera successfully detects the movements of facial expression. The study and use of Jones Viola Algorithm and with help of Haar classifier in successfully extracting the facial images is the objective and its integration of the same with the help of the webcam matches the output

The intended project will take the facial images of the drivers face of measure various facial parameter, nodes on faces thereby working accurately serving its purpose to the fullest of detecting drowsiness. Python and OpenCV have been popular choices among the developers due to its vast library.

AI is currently on the boom. No doubt the future will belong to this domain. A lot of research has been done in the past and multiple advancement has been done in medical applications, fraud detection and self-automated car driving. This model will be an example of the superiority of AI and its branches in information technology.

The project takes the snaps of human faces measures on various inputs of eye-blinking rate, delay in eye blinking and other facial nodes for yawning and head distraction to gather images, match them with already available in database and alert the driver from mishap. The algorithm used are jones viola algorithm, Haar classifier.

The study is expected to apply the concepts and techniques of Artificial intelligence in building efficient sleep drowsy detection system. It's no doubt that the field of AI vast and the level of complexities it has in understanding neural networks deep learning, facial algorithm will offer an proper base for learning among the members.

Understanding of the various algorithm namely jones Viola algorithm popularly in building the system is of prime importance. In addition to that building and system for is compatibility over different devices and its integration in smart phones is of another concern which can be further incorporated or be an interesting base for future research.

IV. CONCLUSION

The algorithm presented in this paper forms the basis of many 'Real Time System' which gives highly accurate results in less time. This system was built keeping in mind every possible behavior of the driver and also the need for a quick computed result which is accurately achieved by the Viola Jones algorithm which is the backbone of Drowsy Detector Algorithm. In the current system when the multiple faces are detected the results are inaccurate. The optimum result would be, if multiple faces are detected then only the driver's face should get detected. This could be achieved using the Depth Calculations wherein the area of the rectangles of multiple faces are calculated and the largest area is selected since in most of the cases the driver's face would be ahead and all the extraneous faces would be behind the driver.

References

[1] Eddie Galrazza,Franklin S Silva"Realtime driverness detection using system of Human Computer interaction implemented in smart phones "2015.
[2] Ankit Jaiswal Mrs Rachana Modi "Face Eye detection techniques in Driver drowsiness system "2014.
[3] Arun Sahayadhas ,kenan Sundarraj , Murugappan and Murrugapan light"Detecting drowsiness based on sensors "2013.

Heart Disease Prediction using Machine Learning and Artificial Intelligence

Darpan Mehra, Varun Mishra, Shridhar Kamble
Information Technology, Thakur College of Engineering and Technology/ Mumbai University, India

Abstract – **Heart related diseases or Cardiovascular Diseases are the main reason for a huge number of deaths in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases.**

Data analytics is useful for prediction from more information and it helps medical centre to predict of various disease. Huge amount of patient related data is maintained on monthly basis. The stored data can be useful for source of predicting the occurrence of future disease. Some of the data mining and machine learning techniques are used to predict the heart disease, such as Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM).

Prediction and diagnosing of heart disease become a challenging factor faced by doctors and hospitals both in India and overseas. The main objective of this research project is predicting the heart disease of a patient using machine learning algorithms.

I. INTRODUCTION

Heart diseases have emerged as one of the most prominent cause of death all around the world. According to World Health Organisation, heart related diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the leading cause of mortality [1]. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15, 2017. Thus, feasible and accurate prediction of heart related diseases is very important. Current approaches to predict cardiovascular risk fail to identify many people who would benefit from preventive treatment, while others receive unnecessary intervention. Today numerous doctor's facilities introduced some kind of quiet's data frameworks to manage their social insurance or patient information. These data frameworks commonly produce a lot of information which can be in distinctive organization like numbers, content, diagrams and pictures yet sadly, this database that contains rich data is once in a while utilized for clinical choice making. There is a lot of data put away in stores that can be utilized viably to guide a medical practitioners in decision making in human services.

II. II. MATERIAL AND METHODS

This project was carried out using data available on the website of University of California – Irvine to train our models.**Study Duration:** September 2018 to February 2019.

Number of Attributes: 75

Number of Significant Attributes: 13

1. #3 (age)
2. #4 (sex)
3. #9 (cp) (chest pain type)
4. #10 (trestbps) (resting blood pressure (in mm Hg on admission to the hospital)
5. #12 (chol) (serum cholesterol in mg/dl)

6. 6. #16 (fbs) (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. #19 (restecg) (resting electrocardiographic results)
8. #32 (thalach) (maximum heart rate achieved)
9. #38 (exang) (exercise induced angina (1 = yes; 0 = no))

10. #40 (oldpeak) (depression induced by exercise relative to rest)
11. #41 (slope) (the slope of the peak exercise ST segment)
12. #44 (ca) (number of major vessels (0-3)

colored by fluoroscopy)

13. #51 (thal) (3 = normal; 6 = fixed defect;

7 = reversible defect)

14. #58 (num) (the predicted attribute)

We are using Tensorflow, Numpy, SciPy, and PyTorch; all of which are open source frameworks written in Python.

To avoid overfitting of the model, we would be following a step by step analysis of our deep learning model. The end result of the analysis is achieving accuracy of more than 90% for a cardiovascular disease and stroke prediction.

Algorithms used for analysis:

1) Random Forest Algorithm

Random forest is a concept falling under the general technique of random decision. This algorithm operates by creating a group of decision trees at training time

and outputting the class that represents the mode of classes or the mean prediction of the individual trees. Individual decision trees are generated using a random selection of attributes at each node to determine split. During classification, each tree casts a vote and the most popular class is returned. Using the Random forests, the variance can be reduced by averaging the deep decision trees trained with different parts of the training set. To form Random forests, tree predictors should be integrated in a way that each tree be dependent on the values of a random vector sampled independently and uniformly from all trees in the forest. We use this approach to predict flight delays in our database.

Random Forest is also a popularly supervised machine learning algorithm. This technique can be used for both regression and classification tasks but generally performs better in classification tasks. As the name suggests, Random Forest technique considers multiple decision trees before giving an output. So, it is basically an ensemble of decision trees. This technique is based on the belief that a greater number of trees would converge to the right decision. For classification, it uses a voting system and then decides the class whereas in regression it takes the mean of all the outputs of each of the decision trees. It works well with large datasets with high dimensionality

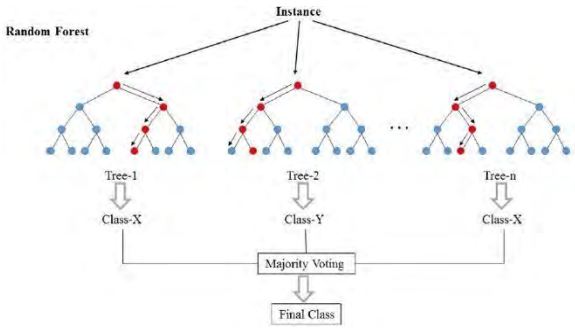


Figure 1: Random Forest Algorithm

2) Support Vector Machine (SVM)

Support vector machines exist in different forms, linear and non-linear. A support vector machine is a supervised classifier. What is usual in this context, two different datasets are involved with SVM, training and a test set. In the ideal situation the classes are linearly separable. In such situation a line can be found, which splits the two classes perfectly. However not only one line splits the dataset perfectly, but a whole bunch of lines do. From these lines the best is selected as the "separating line".

The best line is found by maximizing the distance to the nearest points of both classes in the training set. The maximization of this distance can be converted to an equivalent minimization problem, which is easier to solve. The data points on the maximal margin lines are called the support vectors. Most often datasets are not nicely distributed such that the classes can be separated by a line or higher order function. Real datasets contain random errors or noise which creates a less clean dataset. Although it is possible to create a model that perfectly separates the data, it is not desirable, because such models are over-fitting on the training data.

Overfitting is caused by incorporating the random errors or noise in the model. Therefore, the model is not generic, and makes significantly more errors on other datasets. Creating simpler models keeps the model from over-fitting. The complexity of the model has to be balanced between fitting on the training data and being generic. This can be achieved by allowing models which can make errors. A SVM can make some errors to avoid over-fitting. It tries to minimize the number of errors that will be made. Support vector machines classifiers are applied in many applications. They are very popular in recent research. This popularity is due to the good overall empirical performance. Comparing the naive Bayes and the SVM classifier, the SVM has been applied the most

3) Artificial Neural Network (ANN)

These are used to model/simulate the distribution, functions or mappings among variables as modules of a dynamic system associated with a learning rule or a learning algorithm. The modules here simulate neurons in nervous system and hence ANN collectively refers to the neuron simulators and their synapsis simulating interconnections between these modules in different layers [6]. The defining aspect of an ANN is the function implemented at each neuron and the learning algorithm for the dynamic weights assigned to the interconnections among neurons. What makes ANN stand apart is its ability to simulate human thought process coupled with continuous learning, growth and evolution. Also, it is capable of handling large number of parameters and large set of data with noise and yet achieves high accuracy.

Neural Network is built by stacking together multiple neurons in layers to produce a final output. First layer is the input layer and the last is the output layer. All the layers in between is called hidden layers. Each neuron has an activation function. Some of the popular Activation functions are Sigmoid, ReLU, Tanh etc. The parameters of the network are the weights and biases of each layer. The goal of the neural network is to learn the network parameters such that the predicted outcome is the same as the ground truth. Back-propagation along loss- function is used to learn the network parameters.

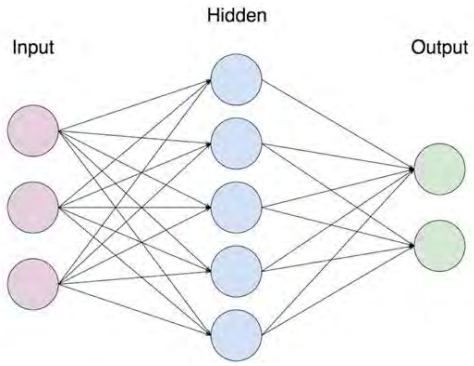


Figure 2: Artificial Neural Network

Sigmoidal Activation Function:

A sigmoid function is real-valued and differentiable, having a non-negative or non- positive first derivative, one local minimum, and one local maximum.
Sigmoid functions are often used in artificial neural networks to introduce nonlinearity in the model.
A neural network element computes a linear combination of its input signals, and applies a sigmoid function to the result. A reason for its popularity in neural networks is because the sigmoid function satisfies a property between the derivative and itself such that it is computationally easy to perform.
The activation function is given by:

1. Bipolar Binary Function:

$$f(x) = \frac{1}{1+e^{-\lambda x}}$$

2. Bipolar Continuous Function:

$$f(x) = \frac{2}{1+e^{-\lambda x}} - 1$$

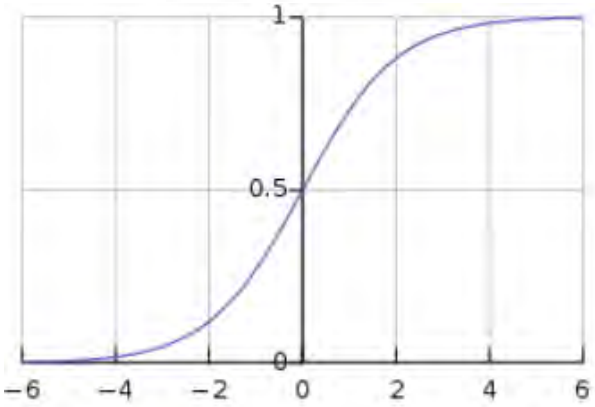


Figure 3: Sigmoidal Activation Function

ML, a branch of Artificial Intelligence, relates the problem of learning from data samples to the general concept of inference. Every learning process consists of two phases: (i) estimation of unknown dependencies in a system from a given dataset and (ii) use of estimated dependencies to predict new outputs of the system.
Figure 4: Proposed Methodology

Figure 4: Proposed Methodology

III. SOFTWARE AND LIBRARIES

Python:

To collect data a web scraper programmed in Python was used. Python's syntax allows programmers to express concepts in fewer lines of codes[13].
Python is intended to be a highly-readable language. It is designed to have an uncluttered visual layout, frequently using English keywords where other languages use punctuation. Furthermore, Python has a smaller number of syntactic exceptions and special cases than C or Pascal. Python uses whitespace indentation.
Why use Python for web scraping and not another thing? Python offers a module called 'urllib2', which has suitable functions to open websites and extract information easily.
Python is used to program the web scraper that is in charge of collecting the weather data for the model.

TensorFlow:

TensorFlow™ is an open source software library for high performance numerical computation. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices. Originally developed by researchers and engineers from the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains. [14]

IV. RESULT

After thorough analysis and synthesis of the training data as well as the test data, we were able to achieve an accuracy of more than 95% in all our three algorithms. On combining the results of these algorithm, a classification accuracy of 97.3%.

In other words, we are able to predict the cardiovascular diseases in 97 out of 100 patients. A report(graph) is shown below for all the algorithms:

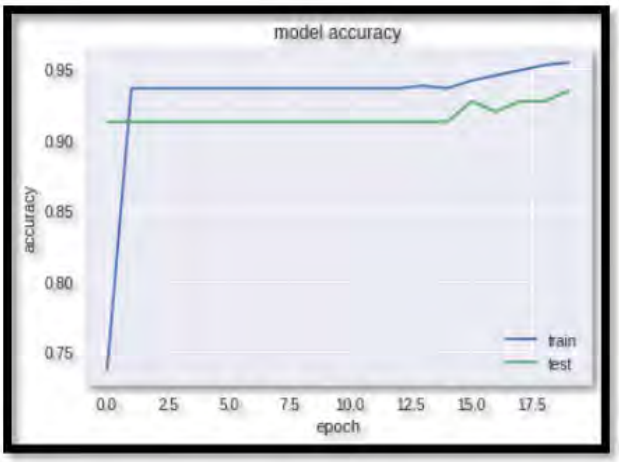


Figure 5: Random Forest Accuracy

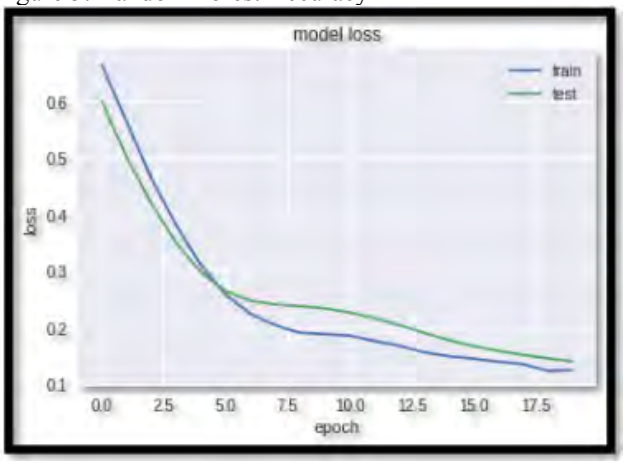


Figure 5: Random Forest Loss

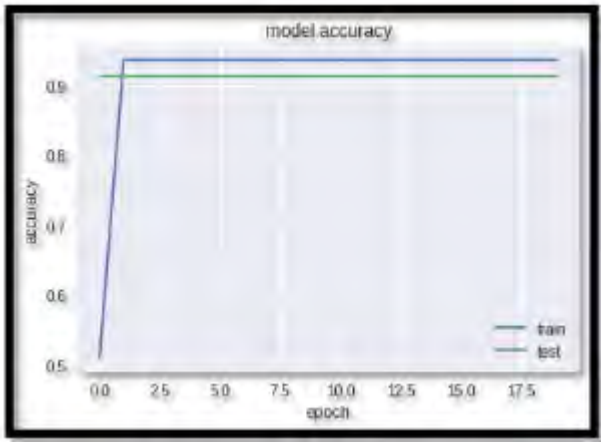


Figure 6: Support Vector Machine Accuracy

Figure 7: Support Vector Machine Accuracy

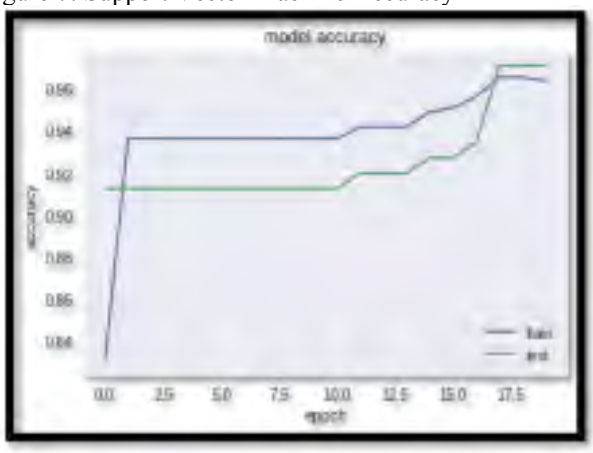


Figure 8: 3-Layer Neural Network Accuracy

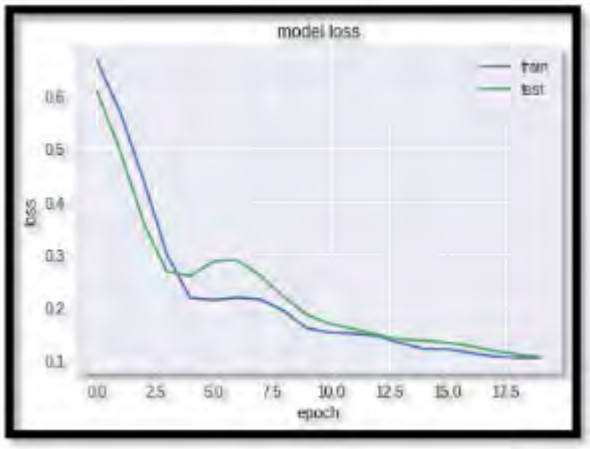


Figure 9: 3-Layer Neural Network Loss

V. CONCLUSION

This project provides the deep insight into machine learning techniques for classification of heart diseases. The role of classifier is crucial in healthcare industry so that the results can be used for predicting the treatment which can be provided to patients. The existing techniques are studied and compared for finding the efficient and accurate systems. Machine learning techniques significantly improves accuracy of cardiovascular risk prediction through whichpatients can be identified during an early stage of disease and can be benefitted by preventive treatment.

It can be concluded that there is a huge scope for machine learning algorithms in predicting cardiovascular diseases or heart related diseases. Each of the above-mentioned algorithms have performed extremely well in some cases but poorly in some other cases.

References

[1] Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, Dimitrios I. Fotiadis, "Machine learning applications in cancer prognosis and prediction", Computation and structural journal, Elsevier, 2014.

[2] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang, "Disease Prediction by Machine Learning Over Big Data from Healthcare Communities", special section on healthcare big data, IEEE access, 2017.

[3] Min Chen, Yujun Ma, Yong Li, Di Wu, Yin Zhang, and Chan-Hyun Youn, "Wearable 2.0: Enabling Human-Cloud Integration in Next Generation Healthcare Systems IEEE Communications Magazine, 2017.

[4] Dr. S. Vijayarani, Mr. S. Dhayanand, "Kidney Disease Prediction Using SVM And ANN Algorithms", International Journal of Computing and Business Research (IJCBR), 2015.

[5] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, Google Brain, "Tensorflow: A System for Large-Scale Machine Learning", 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), 2016.

[6] Jürgen Schmidhuber, "Deep learning in neural networks: An overview", Neural Networks Elsevier, 2014

[7] Yong-Young Kim and Mi-Hye Kim, "What are Software Developers' and Medical Experts' Priorities for Adopting a Healthcare Software Platform?", Indian Journal of Science and Technology, 2016.

[8] https://en.wikipedia.org/wiki/Agile_software_de velopment

[9] https://en.wikipedia.org/wiki/Amazon_Alexa

[10] https://en.wikipedia.org/wiki/Siri

[11] https://en.wikipedia.org/wiki/Google_Assistant

[12] http://agilemanifesto.org/principles.html

[13] https://en.wikipedia.org/wiki/Python_(program ming_language)

[14] https://www.tensorflow.org

Public Address System

Krupal Vora, Neetala Parkar, Dhwani Desai, Dr. Rajesh S. Bansode
Information Technology, Thakur College of Engineering and Technology, India

Abstract – **The Public Address System is used to notify or address a group of people in a particular area over the speakers. The general use of the Public Address System is in the areas such as factories, fire departments, railways and many more. The current PA system contains a microphone which is connected to multiple speakers in the different locations. This current system is not very effective in the railways due to cacophony present in the environment of a railway station and the commuters are not properly notified. This project would help to minimize the problem by notifying the people through their personal devices. The existing systems that applied wireless public address system were unable to solve several technical challenges such as undiscoverable network, range limitation and jamming signals. The proposed system will be able to conquer the above limitations in an efficient manner. Public Address systems are being used at almost all the railway station in India. Although the world is becoming increasingly networked with each passing day, the PA system at stations remains an isolated one. The information is to be given to people for train availability and any emergency arising message on every passenger's phone in rail industries. A better approach is adopted to build a system which would help to overcome the problems of the current system. The new system would be able to send textual notifications to user's present on a particular railway terminal through application notifications process to user's personal device such as, the mobile phone.**

Keywords: MFCC Algorithm, Gaussian Mixture Model (GMM), notifications, text messages, public address system.

I. INTRODUCTION

A public address system (PA system) is an electronic distribution system using a microphone, amplifier and loudspeakers and sound amplification , used to allow a person to address a large public, for example for announcements at noisy and large air or rail terminals or at a sports stadium. The Railway network in Mumbai is 319kms long. There are five routes. They are two western railways, two Central railways and a harbour railway, which assists
6.6million to people travel through Mumbai. According to records, Mumbai Railway has the highest density of passengers in the world. To communicate information to a huge crowd on the railway a PA system is used. A new approach can be implemented by text based message system, using Mel-Frequency Cepstral Coefficients (MFCC) algorithm and Gaussian Mixture Model (GMM).

II. II. Material And Methods

1. Requirements

MATLAB
Android Studios
Android Device
Speech to text recognition software (MFCC Algorithm)

Server
Microphone

2. Methodology

2.1. MFCC

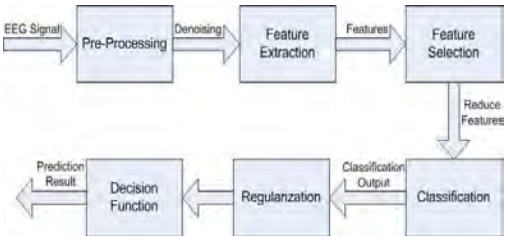


Fig 1: MFCC Algorithm Flow

The Mel-Frequency Cepstral Coefficients (MFCC) is one of the most prevalent and easy method to extract the spectral features from a human voice. The MFCC algorithm is widely popular among the speech recognition systems. The main features of the MFCC
include pre-processing or pre-emphasis, framing, feature extraction etc. The pre-emphasis method is needed because the components with high frequency have small amplitude as compared to the low frequency components. The pre-emphasis is the first step in the MFCC which helps to boosts these high frequencies for quality output. The input in the MFCC is a human voice i.e. analog signal, the analog signal is converted to digital signal to create samples. The framing method in the MFCC is used to create frames and acquire information from the samples. The feature extraction is a method to extract the features from the voice samples to create a cluster using the Gaussian Mixture Model (GMM).[1]

2.2. Gaussian Mixture Model (GMM)

The Gaussian Mixture Model (GMM) is a parametric probability density function which is represented as a weighted sum of Gaussian component densities. The GMM is a parametric model of probability distribution function of measuring features in a biometric system. The main feature of the GMM is used as a classifier to compare the different features extracted from the samples by the MFCC and create clusters of features with the Maximum Likelihood. The GMM is represented by Gaussian distribution and each Gaussian distribution is calculated by its mean, variance and weight of the Gaussian distribution.

The GMM uses the K-Mean algorithm for the creation of clusters.[1]

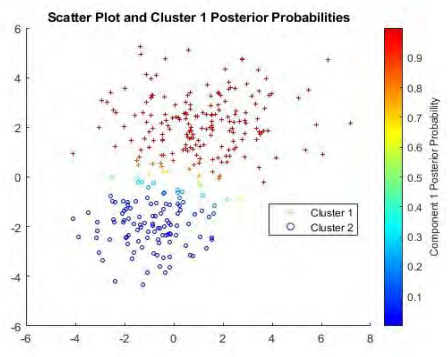


Fig 2: Scatter Plot and Cluster

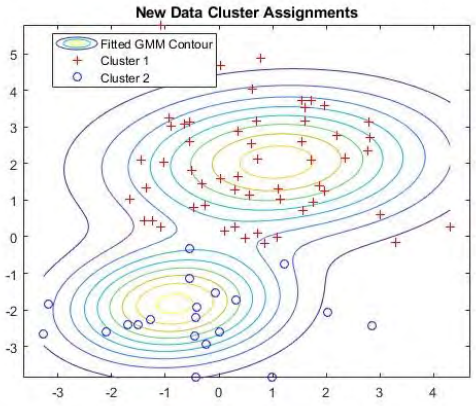


Fig 3: Cluster formation using GMM

2.3 System Flow

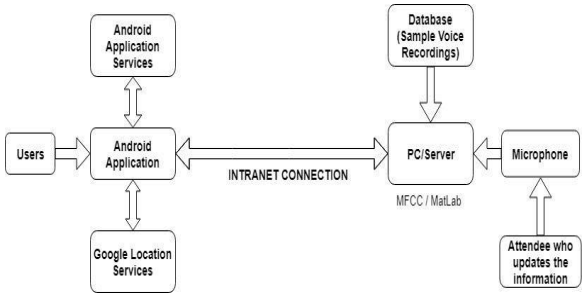


Fig 4: Block Diagram

- Referring to fig 4, the block diagram explains the working of text based public address system.
- In this approach we use text notification for addressing the passengers on the railway platform. According to our problem definition and objectives.
- The railway attendant will announce the update using a microphone which will pass through the MFCC Algorithm and Gaussian Mixture Model (GMM) for speech to text conversion.
- The registered passengers would receive the update via application notification, if the users are present in the mentioned range then and only then will they receive the messages which will be further tracked by the Google location API services installed in the android application.
- Android application helps in tracking the location of the passengers and thus a message will be sent on his/her personal device via text notifications.

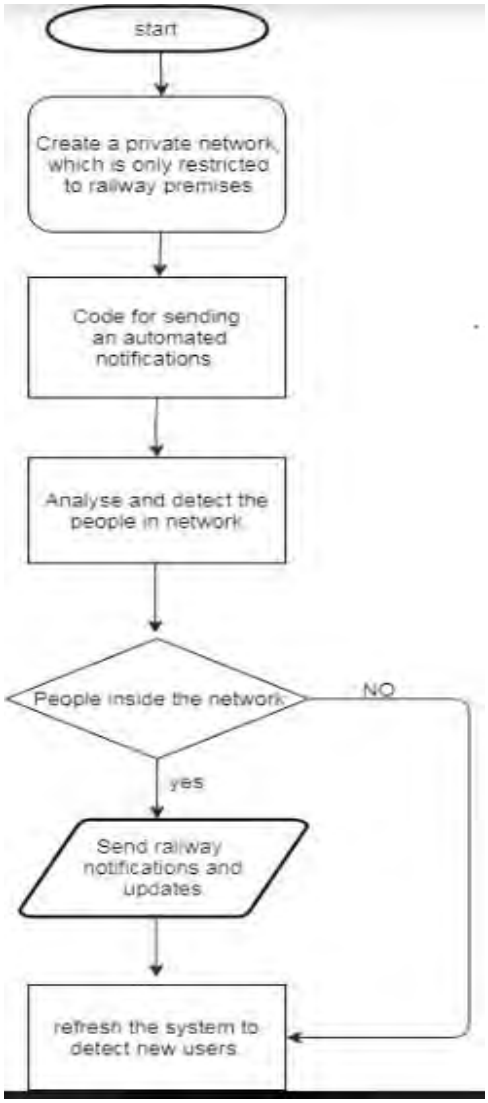


Fig 5: Flowchart

III. III. Result

A) Output

The new approach to public address system will provide the user/people travelling by the railway with satisfactory delivery of information. The system will convey the information to the personal devices of the

users through text messages. The main purpose of our system is to provide efficient means of communication medium to the users. People within the range of the Railway terminal will acquire the relevant information related to the terminal.

Fig 6: Announcement for Mumbai Central to Virar

In the above fig 6, the announcer provides a voice input regarding the information of the train. The announcer specifies that the train will be leaving from Mumbai Central to Virar from platform no.1 at 19:10 hours. The local is on time and is a fast train.

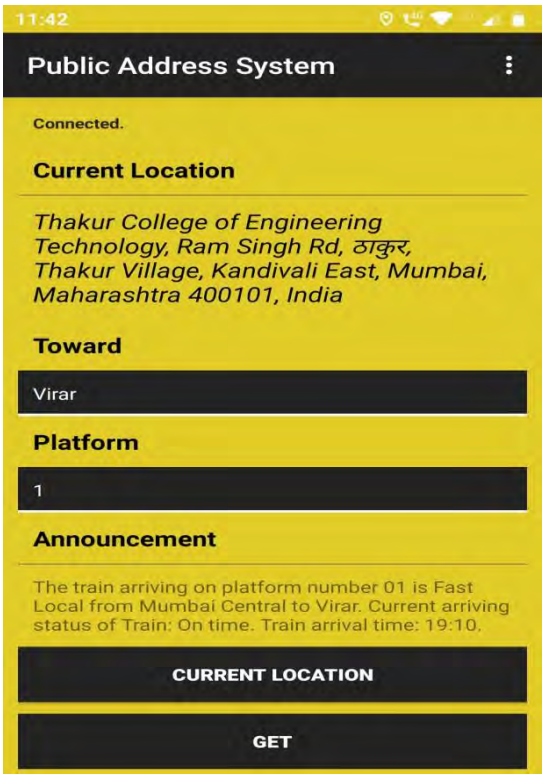


Fig 7: Notification from Mumbai Central to Virar.

The above fig 7, displays the notification for the above announcement for the train leaving from Mumbai Central to Virar from platform number 1. The above figure also displays the current location of the user.

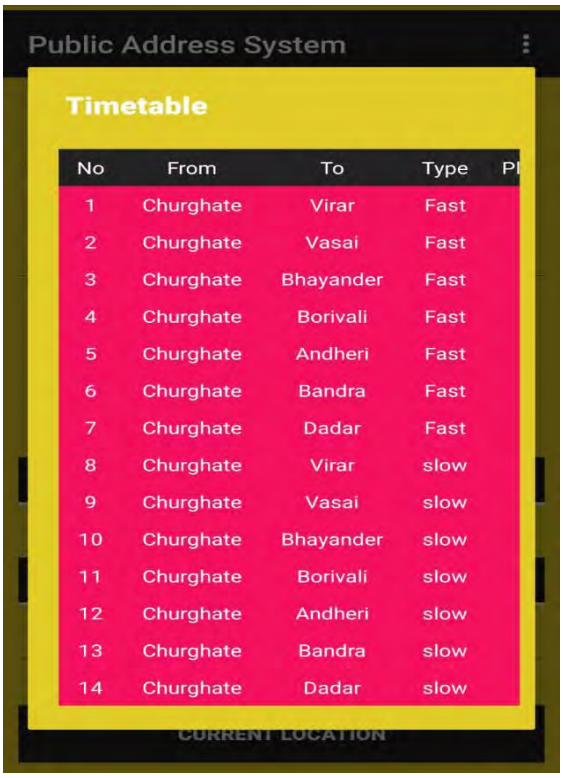


Fig 8: Train Schedule

In the above fig 8, the screenshot is rendering a schedule of the trains. The user can refer to the schedule and decide one particular train accordingly.

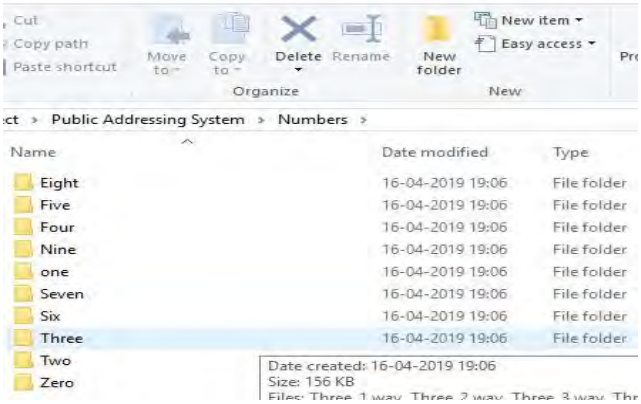


Fig 9: Database for Numbers

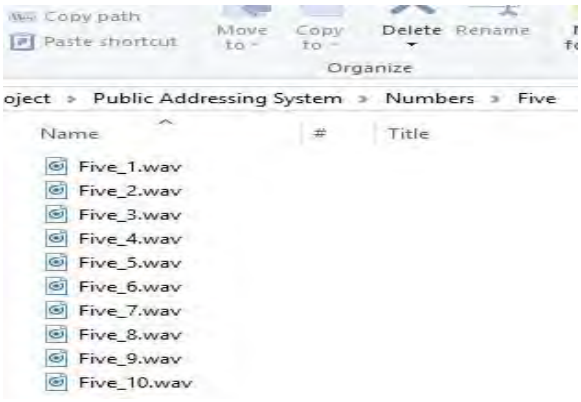


Fig 10: Voice Samples for Number Five

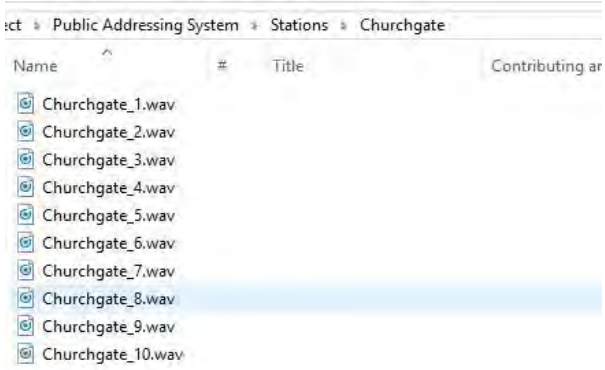


Fig 11: Voice Samples for Churchgate

Voice samples are recorded before the execution. The voice samples are trained by using MFCC Algorithm. The input voice is then compared with the configured voice samples. The algorithm then gives an output in a proper textual format.

Fig 12: Google Location

B) Outcomes

The proposed system will help enhance the current Public Address System for the Railways.

This system will create an efficient communication medium for the Railways to address scores of people.

C) Discussion of Results:

From the literature survey conducted and resulting research gaps acquisitions, the planned project tries to remove all the flaws and tries to provide an effective public address system to convey information to huge crowd. The main goal is to make a user friendly and efficient system for the Ministry of Railway.

IV. DISCUSSION

Public Address systems are being used at almost all the railway station in India. Although the world is becoming increasingly networked with each passing day, the PA system at stations remains an isolated one. The information is to be given to people for train availability and any emergency arising message on every passenger's phone in rail industries.

The need to build a system will help to overcome the problems of the current system. This new system would be able to send textual notifications to user's present on a particular railway terminal through SMS gateway process to user's personal device such as the, mobile phone.

The project aims to convert the speech of an announcer into text and send that text to Android smart phones using intranet. Now while using MFCC algorithm we can get a better voice quality and faster computation we also want to recognize what's said by the user. This is done by using trained samples stored in the databases for interpretation. The project can be mainly divided into two sections, one being the functioning of the MFCC algorithm and second being the development of android application. Understanding the android application development: To run Android apps our project already relies on an internet connection using a web based STT is a much better option. Application will consists of the admin server also the client supported API. The admin will update the information that is to be addressed to the crowd. The application will continuously check for any updates on the server relevant to the location. This will enable the application to send notifications to the relevant users.

Understanding the MFCC Algorithm and GMM: MFCC stands for Mel-frequency Cepstral coefficients. It is a voice recognition algorithm. Voice recognition works based on the premise that a person voice exhibits characteristics are unique to
different speaker. The signal can be greatly different due to many factors such as people voice change with time, health condition (e.g. the speaker has a cold), speaking rate and also acoustical noise and variation recording environment via microphone.

The main purpose of the MFCC is to extract the features from the samples and provide it to the Gaussian Mixture Model (GMM) to create clusters. This is done by a series of steps. The first step is the pre-emphasis method which helps to boost the high frequencies components of the speech signal which are lost during the speech

production. The input which is analog signal is converted into digital signal to create samples. The framing method is used to creates frames and acquire information from the samples. The main step of the MFCC is the feature extraction, in this the step the different from the samples are extracted and supplied to the GMM to create clusters.
The GMM is used as a classifier to compare the different features extracted from the samples by the MFCC and create clusters of features with the Maximum Likelihood. The GMM uses the K-Mean algorithm for the creation of clusters.

V. CONCLUSION

The Public Address Systems are used for communicating to scores of people at a same time. The commonly used systems are wired. The expansion of installing heavy wiring is complex and time consuming. This paper we help overcome the current issue and implement a wireless public address system.
On the basis of literature survey and analysis we conclude that new approach stated is more efficient, low cost yet high energy usage. This method is easier and less complex to implement. The new approach can catch a market in future.

References

[1] V. Chauhan, S. Dwivedi, P. Karale, S.M. Potdar, "Speech To Text Coverter Using Gaussian Mixture Model (GMM)", *International Research Journal of Engineering and Technology (160-164)Vol. 05, Issue No.02,* Feb-
2016
[2] Government of India, "Maintenance handbook on public address system-Indian Railway", Maharajpur, Gwalior, December
2012.[3] A. Hatekar, K. Gaedom, S. Wankhede, P. Dhomane, "Implementation of audio signal processing using DSP 320c6713", *Priyadarshini college of engineering*, Vol.5, Issue No.5,May2016.
[4] Kumar, "Networked PA System", Gyandeep,2017.
[5] S.M. George, "ZIGBEE based Public Address Systems", *International Journal of Computer Applications (0975 – 8887) National Conference on Advances in Computing Communication and Application (ACCA-2015)*,2015.
[6] S. Singh1, G. Mushtaq1, N. K. Tiwari1 and A. P. Singh2, "Cognitive Radio with Software Defined Radio and Mimo for Future Generation Wireless Communication*" Journal of Computer Science & Systems Biology,* Volume 8(3) 169-169 (2015) – 166, 2015.
[7] R. Oshana, "DSP software development techniques for embedded and real time systems", Newnes.p.384 ISBN 978-07505-677592. 23April 2012
[8] Jung-Sook Kim, " Development of a Digital Integrated and Minimized Public Address System with Central Control ", *International Journal of Control and Automation* , Vol. 5, No. 3, September, 2012
[9] Bruce Borgerson (November 1, 2003). "Is it P.A. or SR*". Sound & Video Contractor. Prism Business Media.* Archived from the original on May 20, 2015. Retrieved May 19, 2015.
[10]https://en.wikipedia.org/wiki/Public_address_systems

Filtering political sentiment analysis using social media

Purvi sankhe1 Gaurav Jangir2 Suraj Kanojia3 Krunal Kothari4
Information Technology, Thakur College of Engineering & Technology, India

Abstract – **Using Analytics has been as essential study for the purpose of predicting and classifying data. Nowadays analytics is used in every field ranging from Businesses to sports. Similarly using analytics for analysing recent trends and happenings on social media has been popular. Social media has been a powerful platform for people from around the world to connect with each other and to share their views or sentiments. Therefore, gathering sentiments from social media and applying analytics in order to extract knowledge is an interesting domain. Political sentiment of people using social media is important as it helps us in gaining knowledge about political views of people. There are lots of political organizations around the world which are very active on social media and similarly many people are expressing their political sentiment in the form of tweets, posts, videos on social media. Therefore, analysing these data from the social media helps us in predicting, classifying, gaining insights of recent trends in politics. There are lots of approach to such systems which help in analysing the political sentiment such as Naive Bayes Approach which classifies data based on trained model and similarly classifying it on the basis on decisions applicable through Naive Bayes model. But this approach does not give guarantee for accuracy in classifying data. It is also based on conditional probability which makes it more complex. Another approach to analysing such social media data (tweets, posts, videos) is using lexicon-based approach which uses lexicology based on the patterns of words, emphasis of words used and semantic orientation of words or phrases used in social media posts. This approach has increased accuracy but still there are setbacks as it cannot recognise pure intent.**

The main objective of this System is to analyze tweets from twitter using Twitter API and analyzing social media data for political sentiment using neural network-based approach for predicting, classifying and getting insights of sentiment on the social media. This approach is used as it analyses past human behavior which was absent from lexicon-based approach.

I. INTRODUCTION

"Political Sentiment analysis using social media" is system developed for analyzing the political sentiment or view of the people on social media who use social media regularly to express their political feelings. Here the tweets from twitter is first gathered/extracted, both current tweets and historical using Twitter API. These tweets are gathered based on search criteria lie date, time, keyword, age. Once the data is searched and collected from Twitter API, an Excel file (.CSV file) is generated. Once the CSV file is generated, the algorithm- Lexicon approach is applied and analytics for the tweets is shown. The overall system has two main modules that are, Twitter API and a python (.py file) which consists of algorithms for analyzing, predicting and providing us with knowledge of political sentiments on the social media.

This software can be aimed for analytics and data scientist to gather the political sentiments using both lexicon-based system and neural network system. Then evaluate the effective reports/analysis or take appropriate decisions required. The system is an application of web analytics and can be effective for political or non-political purposes.

Problem definition:

In the proposed system, we will use a Twitter API through which we will gather both current and historical tweets and will gather the required data with the help of search criteria. Then system will contain a Python file which will convert the collected data from Twitter API into a CSV file. This CSV file will then be fed to another python file which will apply the lexicon algorithm and provide us with the political sentiment analytics. CSV file can be modified by applying labeled sentiment and can be used for training neural network. Features of the project includes

Real-time data extraction from pre-processing
Automatic CSV file creation using python file
Search criterion provided for further analysis
Lexicon based approach based on lexicology
Accuracy for prediction through neural network model
Classification provided by model

Background:

A lot of data analysis is being done these days in various sectors like retail industry, corporate sectors in order to increase the business value. These types of analytics maybe misused by political parties to generate curated post according to user sentiments and spread false information.

To prevent this type of misuse, political posts which are unethical or false can be prevented by our system to be spread and reach our audience.

But analyzing political sentiments through social media is kind of difficult as many political posts, tweets are influential in some or the other way, hence it becomes difficult to analyze the exact political sentiment.

II. MATERIALS & METHODS

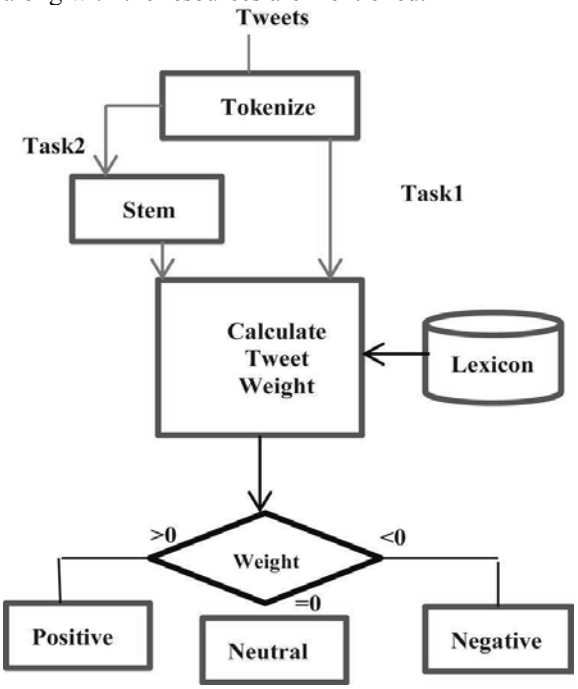
tools. Hence, this project is economically feasible. No heavy Hardware requirement or paid software is in use.

Technical feasibility**:**

This project was based on lexicography and uses Twitter API which is already available in the form of library. Neural Network based approach uses machine learning algorithm requires complex algorithms for accuracy but is feasible from technical point of view.

Agile

Agile methodology mainly focuses on the Operational feasibility**:**
This project requires initial infrastructure, IDEs incremental and the iterative development of the different project phases. It controls and determines how the various stages are implemented and what kind of environment is involved. The project is first divided into different phases, and then the responsibility of developing these phases is given in the hands of the various developers. At every stage of the project development it is continuously monitored and taken care of. These developers developing the phases have particular amount of time period and tasks involved with them. To remain in sync with the other members of the team they have to complete their task within the stipulated amount of time.
The meetings however are conducted on a regular basis where all the members of the team have the right to put forth their ideas along with any requirement or concern. The product backlogs are discussed in these meetings where the requirements of the project to be completed along with the resources are mentioned.



Feasibility Study
Economic feasibility**:**
This project uses all Open Source Software and and tools for initial phase. Though the setup requires integration phase activities but once the setup is done, it is easily operational and is feasible.

Social feasibility:
Generally, this project is based on social media itself, hence social feasibility is a must requirement. The people and political organizations must adopt this system and should find it easy to use our system. Our system is simple in the sense that it uses simple interface.

Stakeholders
The stakeholders include follows: -
Political Organizations.
Social media users
analysts and data scientists

Project Scope
Following would be included in the scope: -
Provide an easy access to social media posts.
Provide better analysis system.
Increase the proper utilization of data for better functionality.
Reduce the maintenance cost of operations.
Improve the accuracy of prediction.
Reduce total cost of the processes.

Requirements and Purpose of the study
Equipment- Hardware:
CPU: Single Core 2.4 GHZ
RAM: 512 MB
Operating System: Windows 10
Hard Drive: 1 GB
Network: Wi-Fi/BroadbandSoftware:
Twitter API library
1.Python IDE
2.CMD Console
3.Python 2 and above

Assumptions and Risk
Assumptions:
- Management will ensure that project team members are available as needed to complete
- project tasks and objectives.
- The employees are willing to change business operations to take advantage of the functionality offered by the new improved technology.
- The core committee will participate in the timely execution of the Project Plan (i.e., timely approval cycles and meeting when required).
- Failure to identify changes to draft deliverables within the time specified in the project timeline will result in project delays.
- Project team members will adhere to the Communications Plan.

Constraints:
Due to the nature of law enforcement,
Resource availability is inconsistent

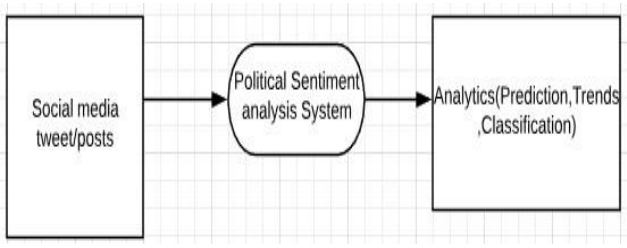
Risks:

Accurate identification of the systemrequirements, interpretation and design depend on the patience and commitment of the people and staff in focus.

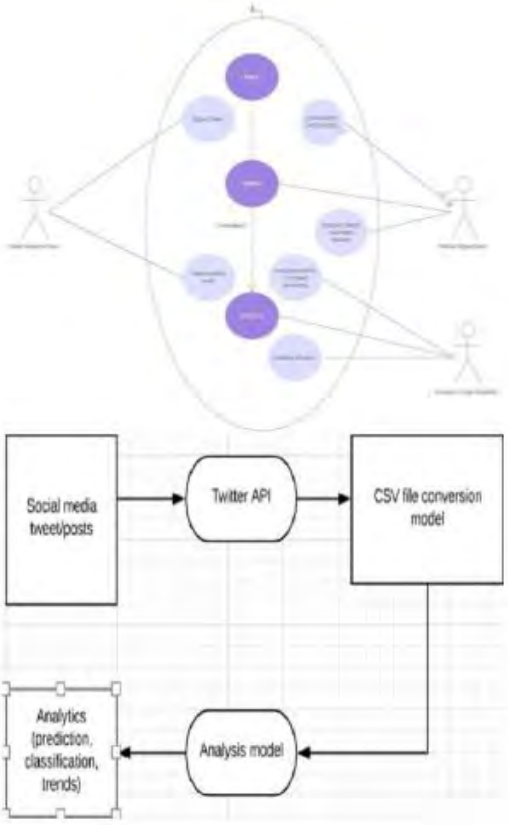
Use Case Diagram

III. RESULT

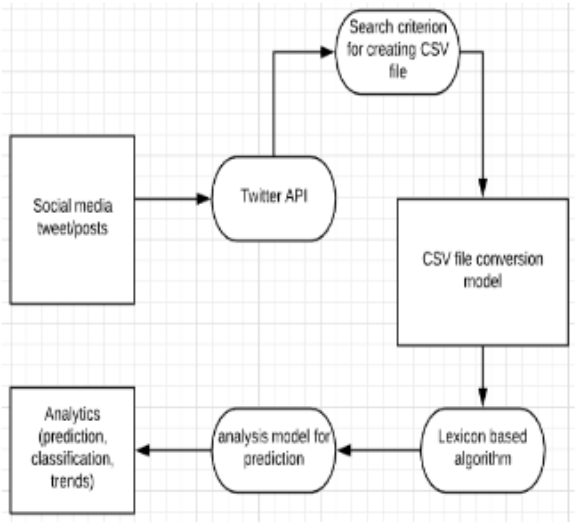
DFD
DFD level 0:



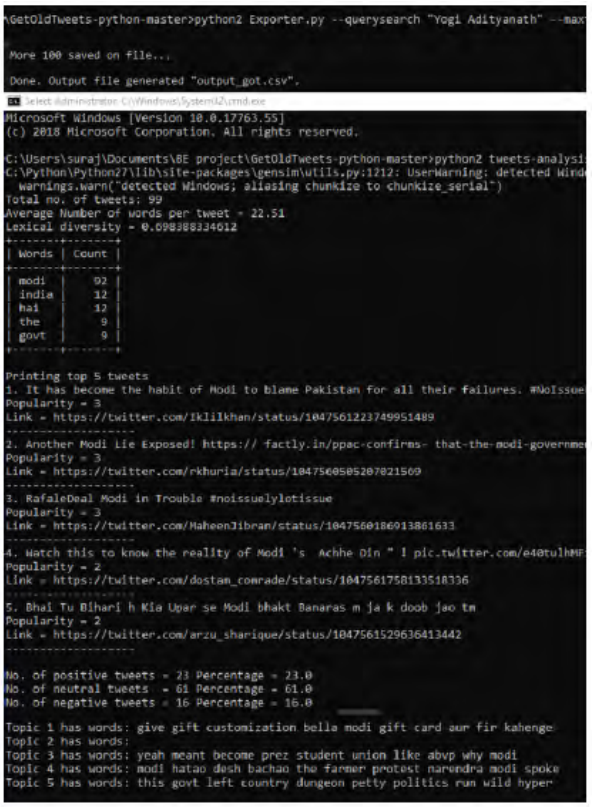
DFD level 1:



DFD level 2:



Output generated using lexicon analysis:



```
\GetOldTweets-python-master>python2 Exporter.py --querysearch "Yogi Adityanath" --max

More 100 saved on file...

Done. Output file generated "output_got.csv".
```

```
Microsoft Windows [Version 10.0.17763.55]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\suraj\Documents\BE project\GetOldTweets-python-master>python2 tweets-analysi
C:\Python\Python27\lib\site-packages\gensim\utils.py:1212: UserWarning: detected Wind
  warnings.warn("detected Windows; aliasing chunkize to chunkize_serial")
Total no. of tweets: 99
Average Number of words per tweet = 22.51
Lexical diversity = 0.098388334612
+-------+-------+
| Words | Count |
+-------+-------+
| modi  |    92 |
| india |    12 |
| hai   |    12 |
| the   |     9 |
| govt  |     9 |
+-------+-------+

Printing top 5 tweets
1. It has become the habit of Modi to blame Pakistan for all their failures. #NoIssue
Popularity = 3
Link = https://twitter.com/IkliIkhan/status/1047561223749951489
--------------------
2. Another Modi Lie Exposed! https:// factly.in/ppac-confirms- that-the-modi-governme
Popularity = 3
Link = https://twitter.com/rkhuria/status/1047560505207021569
--------------------
3. RafaleDeal Modi in Trouble #noissuelylotissue
Popularity = 3
Link = https://twitter.com/MaheenJibran/status/1047560186913861633
--------------------
4. Watch this to know the reality of Modi 's  Achhe Din " I pic.twitter.com/e40tulHMF
Popularity = 2
Link = https://twitter.com/dostam_comrade/status/1047561758133518336
--------------------
5. Bhai Tu Bihari h Kia Upar se Modi bhakt Banaras m ja k doob jao tm
Popularity = 2
Link = https://twitter.com/arzu_sharique/status/1047561529636413442
--------------------

No. of positive tweets = 23 Percentage = 23.0
No. of neutral tweets  = 61 Percentage = 61.0
No. of negative tweets = 16 Percentage = 16.0

Topic 1 has words: give gift customization bella modi gift card aur fir kahenge
Topic 2 has words:
Topic 3 has words: yeah meant become prez student union like abvp why modi
Topic 4 has words: modi hatao desh bachao the farmer protest narendra modi spoke
Topic 5 has words: this govt left country dungeon petty politics run wild hyper
```

IV. V. CONCLUSION

The Political Sentiment analysis using social media is developed using Python IDE, data mining concepts and machine learning concepts. It fully meets the objectives of the system for which it has been developed. The system will reach a steady state where all bugs would have been eliminated as per the planning and Deployment date. The system will be operated at a high level of efficiency and all the political organization and social media users who are active and engaged in politics and users associated with the system understands it. From the tests performed on the new system\design it will improve on the overall performance on the accuracy of social media posts sentiment analysis, thus the system solves the problem that it was intended to solve. Furthermore, computerization of polling system is not just a matter of technological innovation and development; it is a process which involves individual, organization and society in general.

V. DISCUSSION

The project findings show that a dataset of twitter's tweets can be generated using API and this dataset can be used further to analyze and generate model to classify tweets for further filtering of post and safeguard our public interest and intention from political manipulation.

Through this we hope to achieve faster and efficient solution for political sentiment analysis which will reduce the time and complexity of algorithm. The expected result of this project will be to meet political organization's expectation and fulfill all the specified requirements. This project will serve both the political

parties & social media users which will help them analyze the views and sentiments and provide with appropriate decisions to unfold in the future in case of politics.

References:

[1] Journal References:

[1] Prem Melviille, Wojciech gryc, Richard.D. Lawrence, Sentiment analysis of blogs by combining lexical knowledge and Textual Information Vol 9(35), DOI: 10.17485/ijst/2016/v9i35/101807, September 2015.

[2] Taimur Islam, Ataur rahman Bappy, Tanzila Rahman, Filtering political sentiments in social media using Textual Information Vol 9(5), DOI: https://doi.org/10.1109/ICIEV.2016.7760084, May 2016

[2] Web references:

[1]. http://www.w3schools.com

[2]. https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17

[3]. https://dzone.com/articles/applying-nlp-to-tweets-with-python

[4]. https://github.com/Jefferson-Henrique/GetOldTweets-python

The project has a vast scope in future. Project can be updated in near future as and when requirement for the same arises, as it is very flexible in terms of expansion. The following are the future scope for the project:Training the system against large datasets can help in increasing accuracy and scale it to 90%.Election poll result prediction generation

Safety Companion

Harsh Agarwal, Priyank Dave, Tanish Jain, Purvi Sankhe
Thakur College of Engineering and Technology, Mumbai, India
hagarwal754@gmail.com, priyankdave26@gmail.com, jaintanish173@gmail.com, purvi.sankhe@thakureducation.org

Abstract—**Safety companion deals with the development of an information communication technology that can help in protecting vehicles from accidents with the help of the internet of things and a mobile application which works together to protect the driver and other passengers in the vehicles. Along with this, a platform to avail different automobile services as well. The death of the victim can be prevented by removing the causes (pre-accidental) and post accidental by immediately informing the rescue team which will provide the assistance. Thus this system designed here is a system that will bring about the communication to the rescue team to deal with the necessary situation. The system takes various parameters like direction of acceleration, speed, obstacle distances etc. These parameters are used to predict and avoid the accidental situation. The system will send the GPS location of the user with the help of GSM module based on the decision making of the microcontroller (Arduino) and update the rescue team and guardian. Current system doesn't help to avoid pre-accidental causes. The system solves the problem in two parts the one with the mobile application and the other in IoT. Further, the system can be used to track the theft of the vehicle and can also provide an interface for the easy processing and claim of vehicle insurance.**

Keywords— Arduino, GSM, GPS, mobile application, Pothole Detection etc.

I. Introduction

Today, traffic safety is one of the main priorities of the government. Considering the importance of the topic, identifying the factors of road accidents has become the main aim to reduce the damage caused by the road accidents. Road safety means development and management of the roads, provision of safer vehicles, and a comprehensive response to accidents. The costs of fatalities and injury due to traffic accidents have a great impact on the society. In recent years, researchers have paid a great attention towards determination of the factors that significantly affect the traffic managements and roadways accidents. Thus a system for communication of such mishap is needed to be designed. Hence we are designing a system for communication.

II. Background

Road safety becomes major public health concern when the statistics show how more than 3000 people around the world succumb to death daily due to road traffic injury. The huge economic losses are an economic burden for developing countries. The road data are necessary not only for statistical analysis in setting priority targets but also for in depth study not only for statistical analysis in setting the priority targets but also for the in-depth study in identifying the contributory factors to have a better understanding of the chain of events

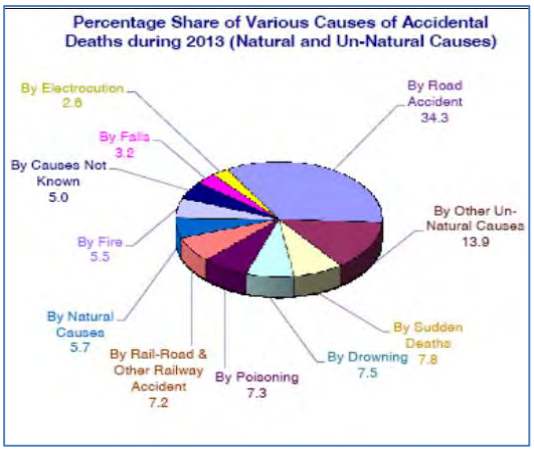


Fig. I: Statistics of major deaths (2003-2013)

The above figure shows the percentage share of various causes of accidental deaths which include Electrocution, falls, and fire, by Rail-Road, Poisoning, Sudden Death, Natural Causes and Un-Natural Causes. The most noticeable one was 'By Road Accidents' i.e. 34.3%.

According to the NCRB (National Crime Records Bureau) report, during the decade 2003-2013, the natural deaths were around 10% whereas the unnatural deaths were the major part. In Unnatural deaths the major cause behind the deaths was the road accident. The road accidents were the major cause of death (of about 42-45%). Maharashtra has reported the maximum deaths due to road accidents (15.7% of the overall accidental deaths).

Road accidents has increased due to surge in the motorization both due to increase in population and increased vehicular penetration.

Further internal mining in the data was done and major reasons of the road accidents were found. They include: over-speeding, drunken driving, distraction to the driver, red light jumping, potholes, bad weather, etc. and certain post accidental reasons for death were lack of communication in time to the ambulance and police department about the crash.

III. Literature Survey

Ancy John, Nishant P.R. et al. [2], proposed a system which is composed of two separate design units: transmitter unit and receiver unit. Security system includes alcohol sensor, eye sensor and smoke sensor. Accident detection system comprises of GPS and GSM in cell phones. As collision occurs, piezoelectric sensor will detect the signal and sends it to ATmega328P

microcontroller. Then, the GPS available in the smart phone will start communicate with the satellite and get the latitude and longitude values and name of place of accident will be send to the previously set phone numbers of relatives, ambulance services etc Accidents due to over speed, drowsy and drunken conditions of the driver are prevented. In heavy traffic zone, speed is controlled automatically without the interference of the driver. Thus we can reduce alcohol and drowsy related road accidents.

Sadad Mahamud, Maliha Monsur, Saniat Rahman Zishan. et al.[3] , explained about a system based on microcontroller Arduino. Here Arduino is our main controlling unit. After receiving the data from the sensors and GPS module it first fetches the data than decode the data and finally execute its operation. Arduino which is the main microcontroller unit is connected with 10 DOF IMU Sensor (accelerometer sensor) etc With an accelerometer you can either get a really "noisy" info output that is responsive or you can get a "clean" output that's sluggish.

Amit.V.Kachavimath1 Nagaraj.C et al.[1], presented that an Accident Alert System using android would be more convenient, relatively secure and utilize fewer resources .In case of the proposed system all the facilities provided are same as in the case of existing system but the proposed system provides all the functionalities in a single device with less cost. This system is developed and implemented to track the vehicle and detect the accident very quickly and easily so that to proceed with the further action items. The Android platform provides several sensors that let you monitor the motion of a device. Two of these sensors accelerometer and gyroscope are always hardware-based and three of these sensors gravity, linear acceleration and rotation vector sensors can be either hardware-based or software-based. Based on the drawbacks of the existing system, we need to design the optimal solution for accident detection using single hardware and software configured device. This helps in identifying and tracking of accident in most effective and less time consuming approach.

Poojitha Shetty, Sachin P C, Supreeth V Kashyap, Venkatesh Madi et al.[5], used Apriori Algorithm for finding Association rule among the items of transitions in each individual dataset based on support and confidence. Further they used Naïve Bayes algorithm for classification. Apriori algorithm require large amount of data storage as it generates candidates table first which consists of analyzing the dataset for each data item(Time complexity=n^n where n is number of items or values in the dataset) hence it takes a lot of time to generate patterns for a dataset of 2MB.

Ms. Gagandeep Kaur, Er. Harpreet kaur et al.[4], used Regression and Co-relation models used for analysis of the accident data and implemented it using R's IDE Integrated Development Environment (R studio) which is graphical and statistical computing tool on road accidental dataset for analyzation. R is free software, GNU package used for analyzing the datasets to vaticinate the hidden patterns and trends. Regression and Co-relation models used for analysis of the accident data. Neural networks are the better way to predict and analyze things in a better way and in a more accurate way with the help of various parameters.

Priyanka A. Nandurge, Nagraj V. Dharwadhkar et al [6], used data partitioning to overcome heterogeneity of data. The proposed method uses k-means clustering method as the main task of segmentation of road accident data. Further, association rule mining is applied to discover the situations related with the occurrence of the whole data set and the occurrence of clusters recognized by the k-means clustering algorithm. The combined result of k-means clustering and association rule produces major information. Segmentation is used to reduce data heterogeneity using a number of measures such as expert knowledge, but there is no guarantee that this will result in the best segmentation of the group including road accidents. Cluster analysis can helps to segment road accidents data. The main difficulty of clustering algorithm is to estimate the number of clusters. In k-means clustering, value of k must be given by the user which is one of the limitations of this algorithm. If the value of k is incorrect then it may lead to incorrect clustering results.

IV. PROBLEM DEFINITION

The usage of vehicle is rapidly increasing and at the same rate the occurrence of accidents has also been increased. Accident due to some factors such as potholes can be prevented and incase of accident in time rescue of the victim can be done. The objective of this project is to build a system that detects potholes and helps driver to avoid them and in case of accident the necessary information of the accident is send to the rescue team and the guardian of the victim, so that necessary actions can be taken in time. Also to provide an application that can help the driver to avail the necessary services such as mechanics, fuel, safety etc. in case of emergency.

Phase 1: It deals with the deep study of the existing systems, by analyzing the previously available research papers and technologies.

Phase 2: The result of the literature survey and analysis based on various algorithms using previously available data on road accidents in various regions.

Phase3: The parameters that are the major reason for the accidents are detected and analyzed used to produce the design of the system. Planning the hardware and software to be used

Phase 4: implementation of the design using technologies planned in the above phases

Phase5: Testing and implementation of the software using the tools and of hardware using remote controlled car.

Phase 6: deployment of the prototype in the cars and to the automobile companies.

V. METHODOLOGY

The project involves use of hardware hence waterfall model is the best suited model to be followed for the development of the system.

Waterfall Model- Classical waterfall model divides the life cycle into a set of phases. This model considers that one phase can be started after completion of the previous phase. That is the output of one phase will be the input to the next phase. Thus the development process can be considered as a sequential flow in the waterfall. Here the phases do not overlap with each other.

- Requirement Gathering and analysis − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.
- System Design − the requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.
- Implementation − with inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.
- Integration and Testing − All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.
- Deployment of system − Once the functional and non- functional testing is done; the product is deployed in the customer environment or released into the market.
- Maintenance − There are some issues which come up in the client environment. To fix those issues, patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

Waterfall model is best suited for the projects where there are less number of requirements and for the hardware project designing.

The basic design of the overall project involves majorly three actors driver, service provider and the Guardian. The User will use the application to gain the services such as fuel, mechanic, etc. The service provider team will take care of the services by allocating the nearest mechanics and fuel provider on the spot where the user needs it. Further wherever in case of accident the Guardian of the driver or the user will be get the alert through SMS along with the location of the victim. Based on the location of the victim the nearest rescue team will allocated at the spot of the accident. The rescues team will take care of all the operation of the taking care of the victims. Below is the UML diagram which explains the complete process and interactions.

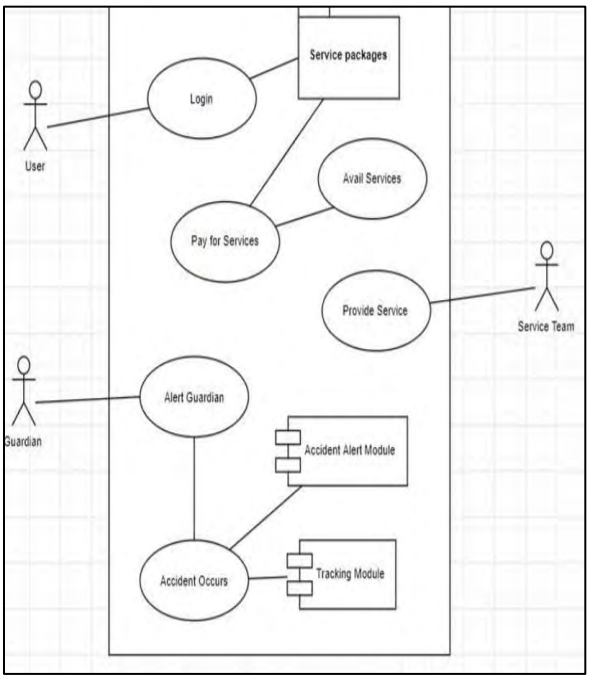


Fig. II: UML Use Case diagram

This UML Use Case diagram provides a use case considering the user/client, its guardian and the service team of the product itself. It deals with features such as login, payment for services, tracking and alerting modules and then alerting the guardian.

The construction of this system is done with the help of Arduino IDE that helps to upload the instructions in the hardware and Android application. Arduino Mega is used to control and adjust the external parameters based on the input parameters. GPS (Global Positioning System) is a small electronic circuit that allows to connect Arduino board to get position and altitude, speed, date and time on UTC.

The block diagram below represents the hardware components such as the ultrasonic sensors, accelerometer, GPS module, Wifi module and the GSM module which altogether work cordially in order to help the client.

The accelerometer senses the tilting of the vehicle. It

detects the abnormal tilt and ultrasonic sensor to detect the obstacles at the back. Thus in case of an abnormal tilt which is consider as a mishap has happened with the passengers and alert to the rescue team and the guardian of the passenger will be generated and necessary services or help will be provided

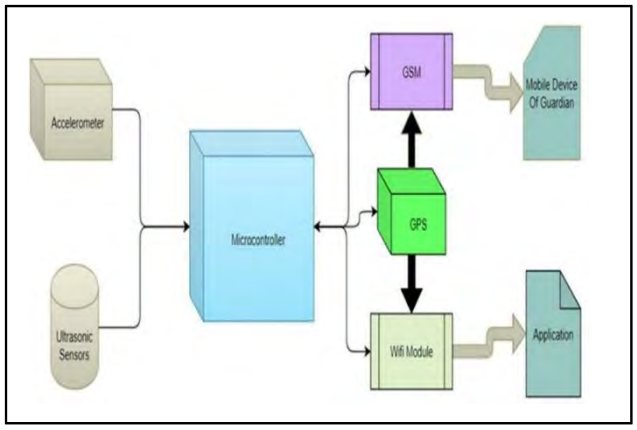


Fig. III: Block diagram of hardware module

The location will be shared to the rescue team and also to the guardian of the victim. GSM chip is used to establish a communication between mobile device or computing machine and a GSM or GPRS system. Accelerometer is a electromechanical device that will measure acceleration forces which may be static or dynamic. Ultrasonic sensor is used to measure distance to a vehicle by using sound waves. It measures distance by sending out a sound wave at a specific frequency and listening for that sound wave to bounce back.

Pothole detection can be done using ultasonic senor, the ultrasonic sensor will be positioned in alignment with the tyres of the car.

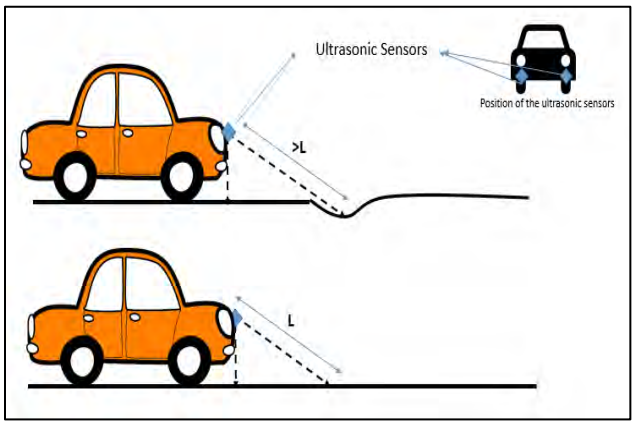


Fig. IV: Pothole detection

The principle on which the potholes are detected here is whenever there is a normal road, the distance calculated by the utlrasonic sensor remains constant whereas whenever there will be a pothole on the right or left tyre the alert of the same before the potholes appear to the driver.

VI. IMPLEMENTATION

The project implementation deals with the development of the application and the hardware implementation. The UI design of the project and the hardware implementation images are shown below in the figures. The UI consists of different tabs that allows user to interact and get the desired services as per his/her needs.

Fig. V: Homescreen

The Homescreen provides the client with the User Interface through which the client can get in touch with the system with just simple touches.

Fig. VI: Service Screen

This screen provides the user with various services which it can avail during an emergency, once it has logged in to the website.

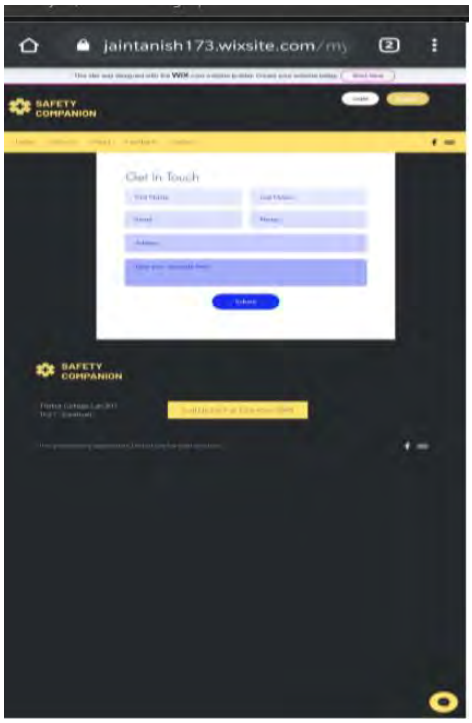
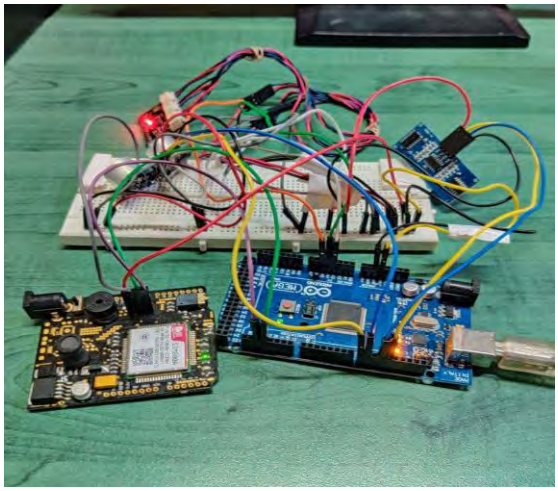


Fig. VII: Registration Form

This form gathers the details of the client/user to which the system can later communicate whenever required along with the personal details.



Fig, VIII: The Hardware of the System

The internal connections of the hardware kit are demonstrated in the above figure which include Ardiuno MEGA, GSM module, Accelerometer etc.

VII. CONCLUSION

Hence we are able to implement 50 percent of the hardware and the software system. Iot system that provides the alerts about the accidents that helps to overcome accidental impact of the people. Thus IoT system has been used to sense the occurrence of the accident depending on the movement of the accelerometer. In case, if the accidents occurs, then the coordinates of the location will sent to the rescue team and to the guardian of the person so that they can reach in time. In addition to this there is a platform provided that user can use to avail different automobile services like mechanical, emergency fuel requirement, cab services, women safety services etc.

VIII. FUTURE SCOPE

In the future, we will further enhance the system so that it will also be able to perform "Potholes detection" which will also help to avoid damage to the vehicle and accidents that can occur because of the potholes. Automation in the car about the acceleration and breaks can brought about to avoid the accidental situations. Moreover, we can add more services in application which will enhance the user experience. Over the next 2 months, the team intends to build a model which will be mounted on the vehicle. The developments of the project in the future will be aimed at careful consideration of further user safety and comfort both on the device as well as the vehicle. Also, we can add a system of a module that will help the user for an easy claim of the vehicular insurance.

References

[1] Amit.V.Kachavimath," Vehicle Accident Detection and Reporting System using Accelerometer ",2015.

[2] Ancy John, Nishant P.R.," Real Time Embedded System For Accident Prevention", 2017

[3] Sadad Mahamud, Maliha Monsur, Saniat Rahman Zishan, "An Arduino Based Accident Prevention And Identification System For Vehicles", 2017

[4] Ms. Gagandeep Kaur, Er. Harpreet kaur, Prediction of the cause of accident and accident prone location on roads using data mining techniques, 2017.

[5] Poojitha Shetty, Sachin P C, Supreeth V Kashyap, Venkatesh Madi, "Analysis of Road Accidents using data mining techniques", in International Research Journal of Engineering and Technology(IRJET), 2017.

[6] Priyanka A. Nandurge, Nagraj V. Dharwadhkar, Analyzing road accident data using machine learning paradigms, 2017.

[7] Miao Chong, Ajith Abraham and Marcin Paprzycki , Traffic accident analysis using machine learning paradigms, 2005.

[8] S. Vasavi, extracting hidden patterns within road accident data using machine learning techniques", 2018

[9] Sannaila Praveen Kumar, M. Prashanti, "Road accident data analysisusing machine learning", 2018

[10] Shrishti Sonal, Soumya Suman, "A framework for analysis of road accidents", 2018.

[11] Road safety and traffic management: Report of the committee Planning Commission, Government of India in February 2007.

[12] Rayle L, Pai M Scenarios for future urbanizations: Carbon dioxide emission from passenger travel in three Indian cities. Transportation Research Record: Journal of the transpotation research board 2193:124-131 (2010).

[13] http://www.deccanchronicle.com/130629/news-current-affairs/article/andhra-pradesh-ranked-3rd-road-accidents last accessed June 29th 2013. [online]

[14] https://towardsdatascience.com/train-your-neurons-on-artificial-neural-networks-featuring-keras-2757f51bb7ed. [online]

[15] https://www.geeksforgeeks.org/data-mining-kdd-process/ [online]

Trend Analysis on Twitter

Sundaresh Naryanan, Abhishek Singh, Kaustubh Shinde, Mr. Rahul Neve
(Information Technology, Thakur College of Engineering and Technology, India)

Abstract – **In social media forums, it is important for public's perception that an organization is well notify and able to provide services that truly meet their needs and interests. Social media activity by citizens is growing exponentially and there is some evidence that certain government agencies are embracing these engagements, and communication tools, albeit at a far slower pace than other societal sectors. The on-going provocation is that the government generally lacks behind when it comes to adaption of technology tools. Social media tools and social media analytics are no anomaly. While this paper focuses on government organizations, it's fair to say all federal organization are playing catch up with commonly employed social media communication platforms. The government is even further behind in the area of social media analytics, the cluster of data from online communication tools and the interpretation of this data to form actionable and relevant information. Solutions to the provocation of this knowledge discovery from the magnitude of social media data must be automated and sophisticated enough to detect sentiment, trends, influencers, and real time issues as they arise. The basis of any social media analytics solution will be the power of its search and discovery knowledge engine.**

Keywords: LDA algorithm, Clustering, Trend Analysis, Social Media Analytics

I. INTRODUCTION

Despite the talks in the market place regarding what social media is, or how it should, impact government operations, the use of information technology to change the way government works and engages with citizens is not new. President Bill Clinton, with the assistance of Vice President Al Gore, initiated what was then called The National Performance Review, which has evolved over time into The National Partnership for generating Government. That endeavour focused on generating greater operational efficiencies, as well as identifying how government could grasp the technology to work more effectively. The overall goal was to cut government costs. This undermines efforts to increase efficiency, effectiveness, and responsiveness. The proliferation of social media throughout American society has put the spotlight on a long standing troubling perception of government – that it is often slow to react and hesitant to move to new solutions and technologies that the general public has already adopted, vetted, and use in a ubiquitous manner. This slow adoptation is often a result of common issues, investigation of potential risks and security concerns, and/or compatibility analysis. Many argue this scrutiny and delay is overbearing and slows the services government agencies provide to the public; others argue it is simply a result of years of entrenched bureaucracy and regulations that lead to slow adaptation to a changing world and dynamic operating environment There are, of course, those who argue that this late-adopter role is one that is necessary to mitigate the risks to the mission of any given agency. In any case, it is clear that while a major part of society has made social media a permanent and complete aspect of their communication strategy, the government, with only a few exceptions, has not yet fully adopted by social media and analytics, and integrated them into how the government conducts its business. What is even unclear is just how government can grasp the myriad of social media tools and channels in order to provide superior service to a better informed and more engaged citizenry.

II. MATERIAL ANDMETHODS

Requirements

Web browser
Windows Operating System
Windows XP, Windows 7,8
MySQL Server

Methodology Used

For developing a project or software it is important to know the flow or which type of software development lifecycle model the system is going to use. Out of various SDLC models there is a need to choose the model which will be perfect for the project and therefore Iterative model has been selected.

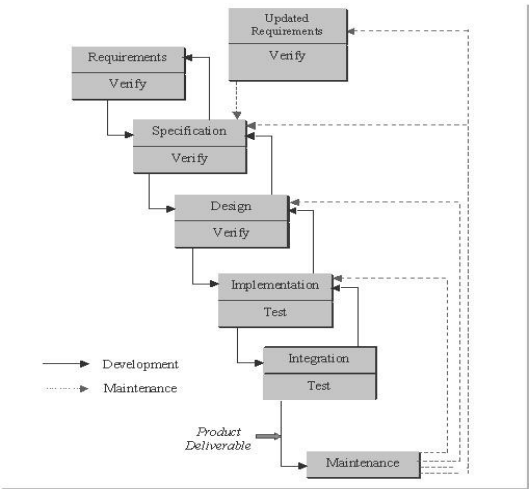
As the name suggest Iterative model is basically an iterative process where the input is processed through various phases and if it doesn't satisfy it is repeated until it becomes satisfactory.

The model consists of five phases namely:PHASE I:

1. Planning: Applying agile methodology for event system event planning for better usage of resources and time with great performance.
2. Analysis: Prepare a detailed analysis on present event system and overcome its limitation in our upcoming system.
3. Design: Integration of data and designing of system event app.
4. Coding: Writing the whole system app code and taking help of an open source.

PHASE II:

1. Testing: Doing various tests on system like unit testing, smoke testing, stress testing, integration testing, regression testing, system testing etc. to check whether the integrated system works as the client desires.
2. Deployment: Conduct beta testing for identifying bugs and errors left out and improvements that can be made. After the testing and approval, deploy the proposed system.

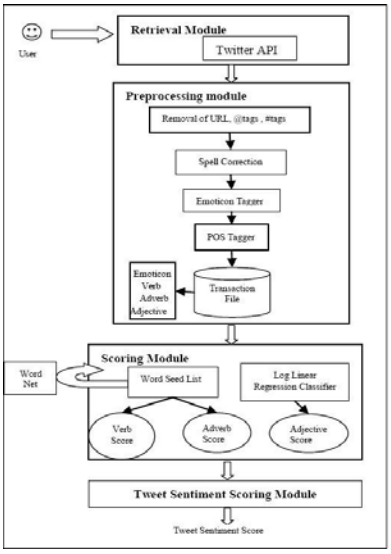


Feasibility Study

Prior developing the product or software it is necessary to know all the factors important for developing a good product. It is important to know the resource requirements, strengths and weakness of the project, etc. All these factors ultimately revolve around two major criteria cost and value, in order to know how much feasible, it is to create a project and how much profit will it return.

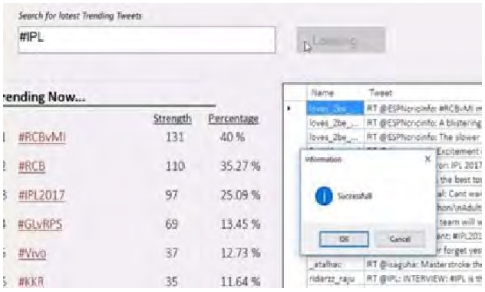
The feasibility of a particular software product can be determined from the following factors:-

1. Technical Feasibility.
2. Operational Feasibility.
3. Economic Feasibility.
4. Legal Feasibility.



III. RESULT

Users can look for social events happening around them through our web application. They can get acknowledged of recent trending events just by entering the keywords. Tweets from all around the world can be accessed with the click of a button **a) Output**



b) Outcomes

The client should be able to connect to the server properly without any problems. The connection establishment between the mobile device and the server should take minimal time. The mobile device should be able receive data from the server uninterruptedly. Information provided by the application should be correct and as per the user's need.

c) Discussion of the results

Connection can be established easily provided that the server is on. The connection with the server takes time as it uses Internet connection. Receiving data from the server takes time. Information coming from the database is correct.

IV. DISCUSSION

Before the introduction of our web application users had to have an access of Television or Newspapers to get aware about the trending events happening around us. To overcome this problem, we introduced a web application that could present you with trending tweets from around the

world with the click of a button. It becomes convenient for users to acknowledge themselves about the events around them through a non-complex application.
Latent Dirichlet Allocation Algorithm has been used to cluster the tweets into different categories.
The application responds quickly while making a decent decision. The application uses LDA algorithm which is efficient for clustering. Irrespective of users location he can access trending tweets around the world. Positive and negative tweets are being clustered to perform
sentimental analysis. Providing security to authorized user using Two
Factor authentication was a one of the major problem.

V. CONCLUSION

In this project we develop a real-time non parametric model to know about the recent trends. It successfully detects the recent trends and it also shows the trends according to the location of a user.

References

[1]. Hsinchun Chen "Exploring extremism and terrorism on the web: The dark web project" 2007
[2]. Amit Pratap Singh, Dr, R. C. Jain "A survey on different phases of web usage for Anomaly user Behavior Investigation" 2007
[3]. Floras Tsai, KapLuk Chan "Detecting cyber security threats in weblogs using probabilistic model" 2007
[4]. Yulei Zhang, Shuo Zeng, Li Fan, Yan Dang, Catherine A. Larson and Hsinchun Chen "Dark web forums portal: Searching and analyzing Jihadist forums" 2009

Innovative Approach for Project Allocation & Monitoring

Divyesh Gohil, YogeshKumar Mutha, Rahul Gupta, Neha Kapadia
IT-Department, Thakur College of Engineering, Kandivali Mumbai , India.
Divyeshgohil97@gmail.com, Yogmutha007@gmail.com,
guptarahulgi97@gmail.com,neha.kapadia@thakureducation.org

Abstract – **Innovative approach for Project Allocation and Monitoring is a web-application that allows the students, guide and admin to track the status of project of students. This project aims at providing for allocation of projects to groups based on their academic performance (like average CGPA of the group). In our project, there are four main modules. In the first module, the system allows the Student, Faculty and the Project coordinators to register themselves successfully in the system. In the second module, the student will login into their particular account and they will do their group formation process and they will set their preferences according to their domain. (The student who will form the group will be by default chosen as the group leader by the system.). In the third module, the allocation script will work, that will be done the by the Admin and the Project will be automatically allocated to a particular group, based on the preferences given by them during the group formation. So the last module goes with the monitoring part, in this the faculty can assign the work to an individual group, and the Group members of that particular group will update the status of the work at the end of the day and everything will be done through the system itself.**
Besides the research, the project also involved the design and development of a system that would help to monitor several projects across several regions. To this end, a web based project monitoring solution is desirable. The web solution has been used to specify and implement the framework to represent all the processes in project monitoring. The use of web technology has been motivated by an increased demand for information sharing, transparency and efficiency in project implementation.

I. INTRODUCTION

Innovative Approach for Project Allocation & Monitoring can be used by organizations or the colleges to make the entire process of allocating the project title online. This project has four roles in this system mainly, Administrator, a Project Guide, Project Coordinator and students. An administrator is the one, who has complete command over the system, he can log into the system and can see each everything which he desires to see (Student details, faculty details, information about a particular group, etc.).
The faculty members will register themselves as the guide and they will give minimum number of the project titles along with their description and their domain expertise. The student will register themselves as an individual and later on the group formation process and preferences selection will be done and individual group will get their own group leader and id by the system.
Now, the allocation script will work, that will be done the by the Admin and the Project will be automatically allocated to a particular group, based on the preferences given by them during the group formation. And the last module goes with the monitoring part, in this the faculty can assign the work to an individual group, and the Group members of that particular group will update the status of the work at the end of the day and everything will be done through the system itself.

II. MATERIAL AND METHODS

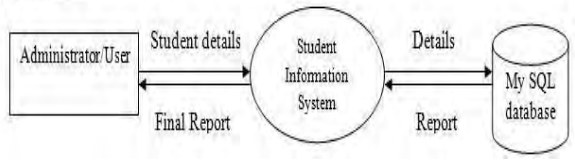
This project application comprises the use of the following resources be it in hardware and software which is discussed as follows:

Software Specifications:

Wamp server
Php dashboard
JavaScript/Html/Css
Code editor (like Sublime, Notepad++)
Browser

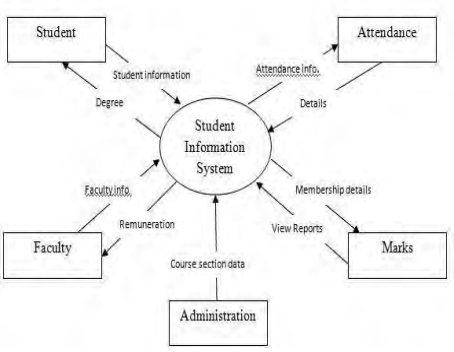
DFD Level 0 is also called a Context Diagram. It's a basic overview of the whole system or process being analyzed or modeled. It's designed to be an at-a-glance view, showing the system as a single high-level process, with its relationship to external entities. It should be easily understood by a wide audience, including stakeholders, business analysts, data analysts and developers.

Level 0 DFD :



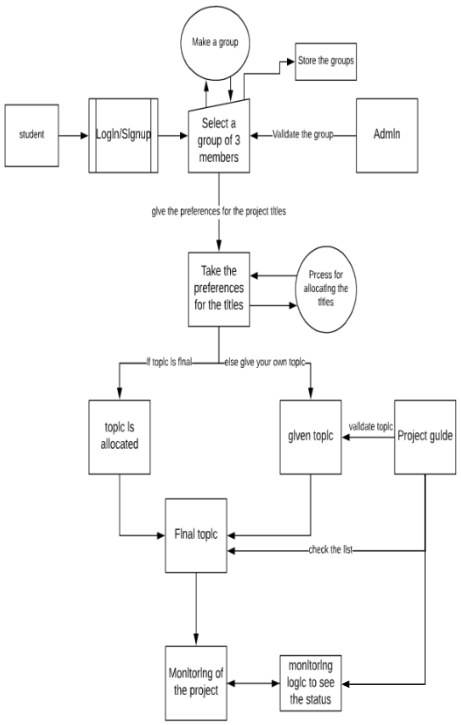
DFD level 0

Level 1 DFD:



DFD level 1

DFD Level 1 provides a more detailed breakout of pieces of the Context Level Diagram. You will highlight the main functions carried out by the system, as you break down the high-level process of the Context Diagram into its subprocesses.



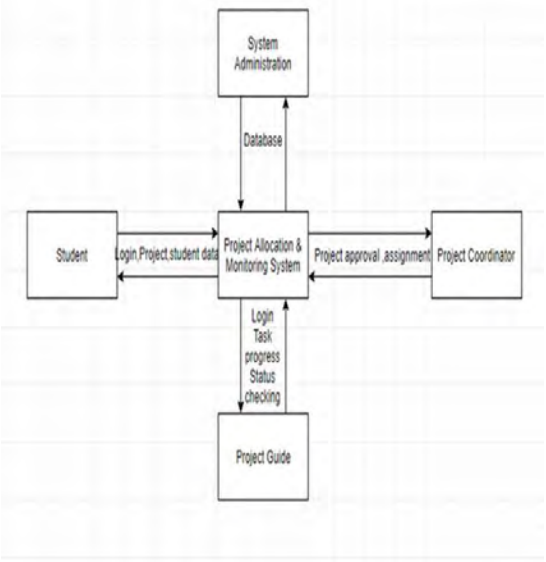
DFD level 2

DFD Level 2 provides a more detailed breakout of pieces of the Context Level Diagram. Now, the studnetWill Register himself over here by filling the basic details of about him. The faculty will register themselves by filling the details about them and giving project ideas along with three topics with their criteria and the description.

The aDmin will be able to look after each and everything and he will run the Allocate script and the Project will than

assigned to a group.

Block Digram :



Block Diagram

This Block Diagram gives a detailed information about the requirements of the Project in terms of the hardware and the software. Various Levels of Data Flow Diagram (DFD) shows the detailed description about the flow of the project. Level 0 describes the Entities involved in the Project, Level

1 Shows the interaction of the Entities with various Other Entities, whereas level 3 shows a detailed flow of the Project, where the student logs in, gives hi preferences for the Project, The faculty registers with the system gives their idea, Admin runs the Allocation script and the project is allocated to the Student. The faculty can assign work to individual group and the students will update from their side which will be tracked by the Faculty.

Procedure Methodology:

Front end for Student and Teacher side will be scripted where student can add his/her project topics & and can see his/her updates regarding the project. Teacher side scripting will be done where he/she monitor the student project development and growth. Admin side scripting will be done where adding of project topics as per message received by the student. Server side scripting will be done where all the data will be stored.

III. RESULT & DISCUSSION

Till now there where many colleges and universities which manually allocate projects assigned to the students. This process is basically carried out for the final year students. It means that the college allocates student group id, project details and a mentor who monitors student's progress. The entire job manually become tedious and time consuming. This project aims at making this job well-based so its saves time and efforts both on students' side as well as on faculty's side. It reduces time of the guide who allocate project to each and every student.

Next thing that motivates us from the project is, there won't be a situation where biasing is done because it will allocate project automatically on the given criteria. The student he or she should fulfill the criteria required to receive the project of their choice. Another motivation is to increase efficiency of current system so that it becomes easy for students as well as faculty to manage and use it.

According to this information project and mentor will be assigned to them. After allocation, students are responsible for fulfilling the milestones set by their mentors. Faculty members, after login, can view the groups under them and can track their progress by checking if deadlines are met proper attendance is met and schedule reports are submitted. A viewer need not log in. He simply can view the gallery aspect of project.

Snap shots of the results are**:**

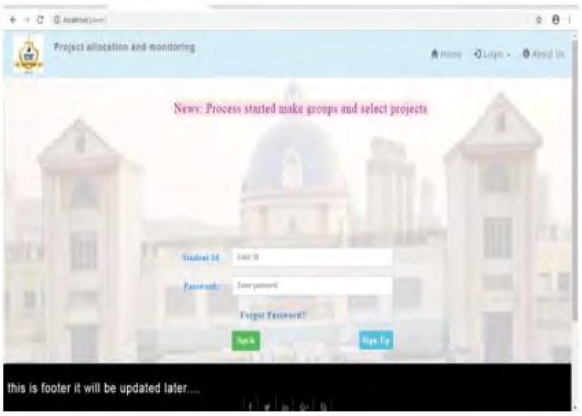


Fig 1 (Login Page For the Students)

Fig 2 (Student Registration)

Fig 3 (Group Formation Process)

Fig 4 (Group Formation and Project Preferences done)

Fig 5 (Faculty Registration)

Fig 6 (Faculty Login)

Fig 7 (Faculty Adding Project Titles)

Fig 8 (Admin Home Page)

Fig 9 (After the Allocation script run by Admin)

Fig 11 (Status tracked by the Guide)

IV. FUTURE SCOPE:

This system is able to manage the database and storing it for a list of student who have registered it.

Easy & safe way for student and faculty registration.

ncrease the proper utilization of time and resources.

Guide can effectively trace the progress of the group and can guide them accordingly.

Reduces the problem of lack of transparency between guide and students.

Student will get the project according to their preferences and the criteria while selecting the Project. projects to groups based on their academic performance (CGPA). The main part of this is the Allocation and monitoring hence the parameters of measurement for the allocation is academic performance like aggregate CGPA of the whole group (cumulative grade point aggregate) will be taken under consideration.

Innovative approach for Project Allocation and Monitoring is a web application that allows the input data to take place securely over the Internet. Finally, the system will able to efficiently monitor and maintain the progress as well as the performance of students on their project. The project guide can easily look upon the task done by the students. Maintain their daily updates and attendance with the faculty. Guide can track more than one group's progress. This entire job manually becomes tedious and time consuming. This project aims at making this job web-based so that it saves time and efforts both on student's side as well as on faculty's side and mapping of attendance can be done.

In our project, there are four main modules. In the first module, the system allows the Student, Faculty and the Project coordinators to register themselves successfully in the system. In the second module, the student will login into their particular account and they will do their group formation process and they will set their preferences according to their domain. (the student who will form the group will be by default chosen as the group leader by the system.).

In the third module, the allocation script will work, that will be done the by the Admin and the Project will be automatically allocated to a particular group, based on the preferences given by them during the group formation. So the last module goes with he monitoring part, in this the faculty can assign the work to an individual group, and the Group members of that particular group will update the status of the work at the end of the day and everything will be done through the system itself.

Features of the project:-

Automatic allocation of project based on academic performance Timely based monitoring of project and attendance To provide last five year project download option for reference To allocate project based on language learned

IV. CONCLUSION

Innovative approach for Project Allocation and Monitoring is a web-application that allows the students, faculty, Project coordinator and admin to track the status of project of students. This project aims at providing for allocation of

Refferences

[1]. Gerasimos Razis (2012) "On-line Project Allocation
[2]. System", Proceedings of the 2009
[3]. IEEE International Conference on management , US Sep2012. ISSN 21681323.
[4]. Omollo, Nicholas , " eProject Tracker " , Washington,united states of America 2011. ISSN: 2357-3451.
[5]. Neha Kapadia, Ashwini Kamat, Priyanka Chaturvedi, Jinal Jain, 2014;"Project Allocation and Status Tracking System",International Journal of Applied Information System(IJAIS)-ISSN: 2249-0868.
[6]. Neha Kapadia, Rajat Bubna, Mitali Jadhav, Arpita joshi,2017,"Project Allocation and monitoring system International Conference on Cloud Computing and Virtualization,283-285.

Diabetes Prediction Model

Dr.Sangeeta, Vhatkar, Nilesh Gupta, Rishabh Jain
Thakur College of Engineering and Technology, Mumbai,India
vhatkarsangeeta@gmail.com, itnileshgupta@gmail.com, jainrish97@gmail.com

Abstract – **Machine learning is a fairly new technology, with new developments coming up faster than most people can learn them. It is a cutting-edge technology in every sense of the word. In such a case, it is pertinent that we handle it with care. It could very well throw up results at an accuracy rate that is infeasible for medical applications. The capabilities required for the project are fairly in the feasible range. Most of the planned feature hinge around processing datasets. Due to the above, a challenge also visible here is the problem of dealing with huge amounts data and processing it thoroughly to get proper results.**
The central parameter of success is the accuracy at which the machine learning algorithm predicts the diseases for which it is trained. It is expected to be above 96% - 98% to make a real difference, as medical diagnosis has very rigorous standards. The secondary parameters are the response and loading time of each modules. A high value of module latency can make any user impatient. Besides these 2, the other parameter is the number of features that work without any bugs. This parameter should be ideally above 90% as well, or it can make users frustrated.

I. INTRODUCTION

While machines and robots may not completely replace doctors and nurses in the hospital, machine/deep learning and artificial intelligence are constantly transforming the healthcare industry. This technology not only has a huge potential, but also offers a whole fleet of applications, which can be implemented in the healthcare sector. The way in which healthcare operates has seen a dramatic change over the last couple of decades, and this change is expected to continue in the years to come.
Evolving from pattern recognition and computational learning theory of artificial intelligence, ML helps in building automated models that explore and analyze the available data sets to identify hidden patterns and make useful predictions, using recognition algorithms [3]. Today, ML is increasingly being implemented in the field of healthcare, especially in image registration, medical image processing, image-guided therapy, computer-aided diagnosis, image database retrieval, image annotation, etc. In this article, we have listed some of the key healthcare applications of ML and the ways in which ML has benefited this industry.
Artificial intelligence presents challenges due the complexity involved in getting the balance between too much and not enough. We may design systems which can process an extremely wide variety of inputs, but we cannot actively
ensure that the AI will respond to it in the way we want it to. It may present us an output which was intended to be for another input, or it may not be able to process it due to the load caused by the heavy processing on the interpreter. Thus, the challenge is not just of input handling or building features, but of efficiency as well.

II. PROPOSED WORK

Analysis of the current market scenario and technological capabilities is essential to the building of a project. Here we analysed various prevalent research papers pertaining to our project as a part of the literature survey. The project aims at creating an intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other hospital staff work efficiently. This also aims at having a basic machine learning capability for disease prediction [3].
Our literature survey was focused on the capabilities
of the choices we made for the core part of the project, namely artificial neural networks and the use of wearable technology for data collection [4]. The initial phases focused on the planning and design of the project. We analysed extant products, customer interests and so on. For analysis, we looked at the current iterations of popular and cutting-edge algorithms used in machine learning, such as artificial neural networks and convolutional neural networks [4]. The analysis was openly available in numerous research papers. The design involved comparing the project with some similar projects.

The working is as follows:

III. LIT ERAT URE SURVEY

Prediction of Diabetes Diagnosis Using Classification Based Data Mining Techniques

In this paper [1], 1) Data mining algorithm renders the last dataset to find useful knowledge. 2) K means is a type of clustering which classifies the data without previous knowledge

Prediction of Onset Diabetes Using Machine learning.

In this paper [2], 1) Machine learning will help us predict early diabetes. 2) Multilayer perceptron is a type of neural network which help us to develop machine learning model.
Detection and Prediction of Diabetes Mellitus Using Back- Propagation Neural Network
In this paper [3] 1) a standard neural network (NN) consists of many simple, connected processors called neurons. 2) Re- cyclic neural networks are the deepest of all neural networks
A novel approach to predict diabetes by Cascading Clustering and Classification.
In this paper [4], 1) Clustering technique can be used to classify diabetes patients according to risk of having diabetes.
2)bagging and random forest can be used to design the model for machine learning

IV. FEASIBILITY STUDY

A .Economic feasibility:

The project relies on open source software, such as python, Scikit-learn, Matplotlib, IPython, Pandas, NumPy and SciPy. The functionality provided by these software's is enough to construct a product capable of usage for healthcare applications.

For the machine's learning and training, we will use dataset from kaggle provided by PIMA. These sets run into thousands of lines, and thus can provide enough learning capability to the software. Due to these 2 being freely available, the project is very low cost and thus feasible from an economic standpoint.

The expected total cost is around 4000 rupees. As such, we do not need any outside funding.

B. Technical

The capabilities required for the project are fairly in the feasible range. The backbone of the project is to create learning model.GUI will be built on using Django framework.
The learning model will be created by using neural networks. Other python packages can also be used for additional features.

C. Operational

Initially, the diabetes dataset is given into the data pre- processing module. The pre-processing module removes the irrelevant features from the diabetes dataset and gives the pre- processed dataset with relevant features to the machine learning algorithm. Then, the machine learning algorithm develops a learning model from the pre-processed dataset.
This learning model is known as knowledge model. The diabetes is predicted for a person's medical report or data using the learning model. Legal
Since the project deals with a sensitive topic (healthcare) with potential for major harm,

Much has been said about the ethical advantages of free software in general, and it is particularly true in a profession in which the sharing of instantly available, accurate information can make the difference between life and death. As medical software begins to offer decision support, risk management, performance rating, and analytic features, physicians should learn to be more careful with its usage.

D.Ethical

While the software is meant to help users with their health, there still is potential for improper usage leading to damage. The software should not be used as a substitute for doctors or professionally provided healthcare. Users must only use it to supplement the same. Much has been said about the ethical advantages of "free" software in general, and it is particularly true in a profession in which the sharing of instantly available, accurate information can make the difference between life and death.

As medical software begins to offer decision support, risk management, performance rating, and analytic features, physicians should not accept black boxes and secret formulas that constrain sharing and intimately affect patient care and reimbursement.

V. V. RESULTS AND DISCUSSION

Table: 5.1 Actual Results

Accuracy: 0.7359				
Confusion Matrix				
[52 28]				
[33 118]				
Classification Report				
Condition	Precision	recall	f1- score	Support
1	0.61	0.65	0.63	80
0	0.81	0.78	0.79	151
Avg./Total	0.74	0.74	0.74	231

VI. CONCLUSION AND FUT URE SCOPE

Currently the project focuses on gathering a structured authentic data. As medical data and

analysis is a sensitive field, the project intends to find an assimilation of the results of various algorithms and compare the accuracy.

Over the next 2 months, the team intends to build a model that best suits the input data. The model and the algorithms used are structured and developed using python as the main programming language and will use libraries like SciKit, NumPY, flask and Rasa for NLP.

The developments of the project in the future are aimed at careful consideration of the disease. This can make a prediction of the diseases by looking at a medical history of a number of patients.

Thereafter, the team intends to build a web framework to create a rich user interface. The comparison between existing results and actual results show a gap in accuracy that has to be closed. However, we were able to achieve more versatility, which was probably the reason that the algorithm was less accurate in the first place.In order todevelop this system, the dataset is collected from the University of California, Irvine(UCI) repository. Different machine learning algorithm namely probabilistic-based naïve Bayes (NB), function-based multilayer perceptron (MLP), decision tree-based random forests (RF) are used to build the machine learning model to carry out the diagnosis of diabetes. Furthermore, the machine learning model is tested with different testing methods such as 10-fold cross validation (FCV), use percentage split with 66% (PS), and use training dataset (UTD) to evaluate the performance of the machine learning model in terms of accuracy.
Different options were taken into consideration to improve the accuracy. So finally by removing outliers, categorizing data, keeping tree depth to 4 we were able to achieve desired accuracy. During this process we figured out few attributes that played an important role . Out of eight attributes Glucose, BMI, Pregnancies, Age and Insulin were the important ones. As per our results and data the other factors like Diabetes Pedigree Function, Skin Thickness and Blood Pressure had negligible effect in determining diabetes

References

[1]. Chunhui, Z., Chengxia, Y., 2015, "Rapid Model Identification for Online Subcutaneous Glucose Concentration Prediction for New Subjects with Type I Diabetes", *IEEE Transactions on Biomedical Engineering*, 62 (5), pp. 1333 –1344

[2]. S.SelvaKumar, .Senthamaraikannan "Prediction of Diabetes Diagnosis Using Classification Based Data Mining Techniques" Volume 12, Number 2 (2017), pp. 183-188.

[3]. Md. Aminul Islam, NusratJahan "Prediction of Onset Diabetes Using Machine learning".

[4]. Miss. Sneha Joshi, Prof. Megha Borse "Detection and Prediction of Diabetes Mellitus Using Back- Propagation Neural Network".

[5]. Palivela Hemant, Thotadara Pushpavathi "A novel approach to predict diabetes by Cascading Clustering and Classification".

[6]. Vinaytosh Mishra, Dr. Cherian Samuel, Prof. S.K. Sharma "USE OF MACHINE LEARNING TO PREDICT THE ONSET OF DIABETES".

Group Messaging Solution

Mr. Abhijit Kushwaha, Mr. Gaurav Gupta, Ms. Nikisha Mistry, Dr. Zahir Aalam
Thakur College of Engineering and Technology, Mumbai, India

Abstract: **In every organization, in order to work collaboratively we need people who come up together and communicate so that they can express their ideas and share best thoughts and have a good team work. In organizations, employees working on same project may be residing in different places of the world to communicate, they need channels which can hold their confidential messages and also protect the messages from networking security threats.**

Currently, we have a number of messaging applications like Slack, Gmail, Whatsapp, etc. Prominently organizations use Gmail for emailing or their own mail sending services for communication. But for mail reception we have a condition that mail receptor should know who has sent him mail. But there can be condition in which someone might spoof, pretending to be someone. For organizations which to segregate authorities as senior and junior and communicate among each other, this kind of platform could be vulnerable. In current systems, hackers could even attach viruses and other malwares to emails. If you want to send messages, you have to manually add the desired recipients.

The project implements application which would provide a way for communicating amongst teams and senior members of the team without hesitating and by implementing this project, members will also be able to put forward their views and points easily in front of the whole group members and senior authorities. Also this project will make the communication effective which will in turn benefit the team and their work. Also threat agents won't be able to send malicious files as this will be only message based and hence more secured. Also it will have a news feed section where latest news about the company will occur which will keep everyone informed about the current scenario and everyone will be on the same page.

The project also aims to identify confidential message sent using this messenger so that the confidentiality remains intact and admin can manage and see which user is using the confidential words also by implementing this, spam and rumor detection can also be implemented.

I. INTRODUCTION

Communication is the first step in every project life cycle. Communication holds key of the entire project because by communication team's form, by communication project working teams do information and requirement gathering of the project. If this step is given the required time and is done sincerely, the project development team will exactly understand what the client wants and they can start planning for the project and complete it without facing client dissatisfaction. For some organizations, the messages are very confidential because they may be passing tender messages that should not leak.

The main aim of our project is to successfully enable organization plan conversations with people based on roles.

The reader of this paper will be able to understand what the objectives of the project are. He will be able to compare the current system and will be able to compare the project and existing applications, overview of the project, importance of the project, background of the project and perspective of the stakeholders and customers.

This project application helps to function efficiently by providing platform for an organization to communicate amongst each other by eliminating the distractions created by other platforms such as Gmail, Facebook, etc. like advertisements, other distractions such as friends email during work and also prevent from sending malicious files. This project also helps to identify any spam, rumor or confidential message by detecting it in its sending state.This project has a news feed section which will keep the employees informed about the latest trends of the organisation.

II. PROPOSED SYSTEM

The proposed system will have all the features we have now thought about such as spam and confidential message detection, no advertisements, affordable, news feed, read aloud messages. Hence this project is meant to construct an application which will ease in communicating amongst themselves and also avoid distractions caused due to advertisements and other spam emails with also a feature to detect confidential or spam or rumor messages in the organization amongst themselves. Also it will have a news feed section which will be operated by the organization itself.

III. METHODOLOGY

We have used Agile methodology for making of this project. In Agile, companies keep iterating the project and testing it during software development life cycle. So this would help us do a lot of tasks simultaneously and develop an application exactly the way the client wants and assure security.

IV. Theory

As mentioned earlier we now understand how important communication for an organization is. So here we have come up with an application that is developed using Node.js.

The application intends to send and receive messages seamlessly with absolutely no delay and chaos. The network security is extremely important for the project. The network will only accept encrypted messages to pass through the network. An unencrypted message can be easily read by any sniffer. This is a kind of passive security attack because receptor of the message or the sender will never come to know that the message that belonged to him has reached him after someone reading it. The application has to also ensure that the attachments do not include a malware. All this is taken into account while developing the application. After the completion of the application in the testing phase, we have tested the application against several possible

security attacks and the application withstood most of them.

The data is stored locally in the application by using MySQL/MongodB database. End-to-End security that let safely exchange private information with each other without worrying about data. In addition to the protection of storage it has lots of modern features such as encryption support. The tools and techniques used by threat actors keep changing and adapting to the conventional security techniques currently in use. Role based DB authorizes the sender receiver and sends the requested resources only if the user has rights to else gives error. Using this system admin will able to get more info like complete logged details of every user including the IP address of each system. Also this messenger will detect any confidential message before sending it and will won't allow to send the message, instead it will alert the admin by informing them about it. It also has a news feed section which will have the current trends managed by the organisation itself.
In this project application the data will be stored periodically in a single machine. No emails required to send the mail hence easing work. Managing is must easier with dynamic Graphical User Interface (GUI). Using AES 128 bit encryption and decryption. A group or team can interact together or individually. Admin manages and assigns the roles of users. It does not have subscription pay scheme and hence saving money and only includes one time fee.

Feasibility Study**:**
Executive Summary

Project is related to Group Messaging Solution. It is similar to whatsapp but GMS will be a desktop application where the user is created by the admin and a unique certificate is generated for each user to identify his/her identity. Users will be assigned into groups i.e. into their respective departments by the admin. Each message will be encrypted by AES 256 bits and will include users certificate at the sender side and decrypted at the receiver side. An OCR technology will be introduced in this which can scan the document and convert the image into text as it is, since in GMS only text is allowed because of safety from Trojans, keylogers, viruses, steganography etc.

- The project maintains two levels of users: -
  - Administrator Level
  - Departmental Level
- Main facilities available in this project are: -
  - Maintaining hierarchy between users.
  - Maintaining logs of chat in the server.
  - Providing secure communication between end users.
  - Providing certificate to each user to maintain their identity.
  - Maintaining backup of data as per user requirements (between mentioned dates).
  - Administrator can only assign the users and provide them their identity.
  - News feed
  - Confidential, spam and rumor Message detection and alerting to admin.
  - Text to speech conversion of message.

Define business problem or opportunity

The Software is for a secure communication in an organization.
It maintains two levels of users: -

- Administrator Level
- Departmental Level

The Software includes: -

- Maintaining logs of chats in database server.
- Providing secure and encrypted communication between end users.
- News feed
- Confidential Message detection

- Purpose

Group Messaging Solution provide following facilities like:-

- ➢ Transparency between admin and users
- ➢ Encrypted Communication
- ➢ Backup of chat
- ➢ News feed
- ➢ Confidential Message detection

Assumptions used in the study

Every user should be:
• Comfortable of working with computer.

Audience impacted

It is designed for industry, start up organisations.

Financial obligation

The financial and the economic questions during the preliminary investigation are verified to estimate the following:

- The cost to conduct a full system investigation.
- The cost of hardware and software for the class of application being considered.
- The benefits in the form of reduced cost.

Recommended action

- Time evaluation is the most important consideration in the development of project.
- The time schedule required for the developed of this project is very important since more development time effect machine time, cost and cause delay in the development of other systems.

Technical Feasibility

- As already many chatting applications are available but when it comes to security GMS is more secure than any other application.

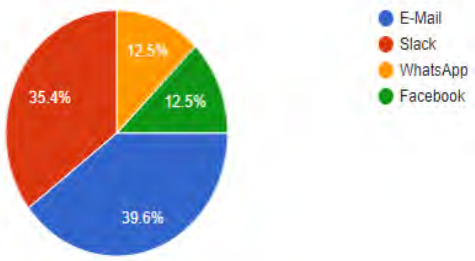
Economic Feasibility

- As GMS is organization specific it will only cost for the database server to keep backup.

Socio-Cultural Feasibility

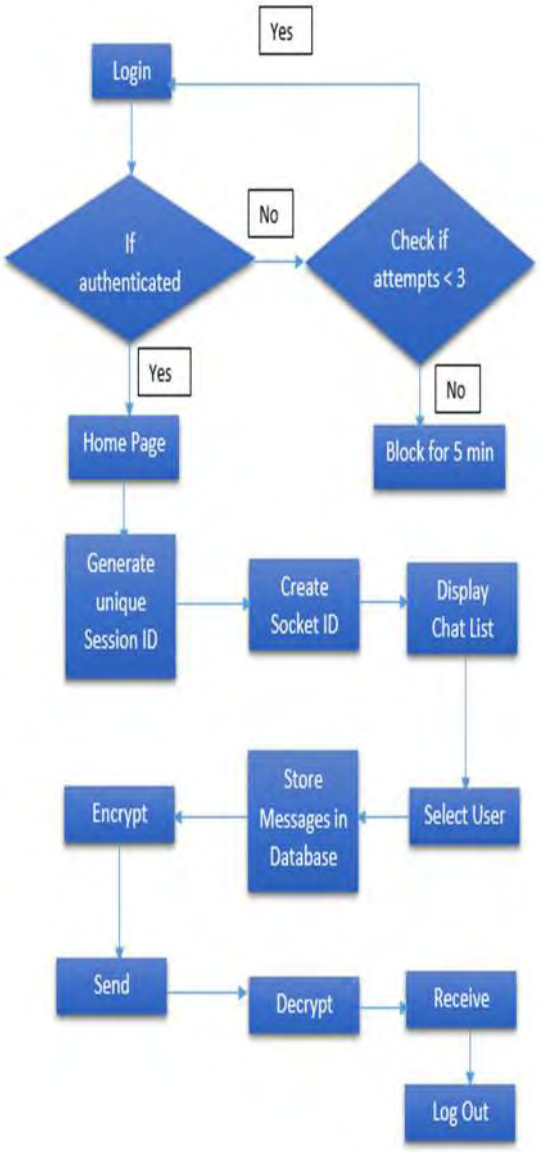
- As GMS is all about group communication there will be complete transparency between everyone.

Research/Technical/Project Work:
The different applications existing in the market today have a lot of features. They are all very easy to use. They are affordable facilities for usage by common man. All the features that the developers claim are indeed worth using. But they cannot be used by business organizations as they have to discuss several formal issues. The applications can have several security threats. We conducted research to study what features are needed by users and which platform they use regularly for communication. The users gave a number of ideas for development of better applications.
The survey can be depicted by the following pie chart:

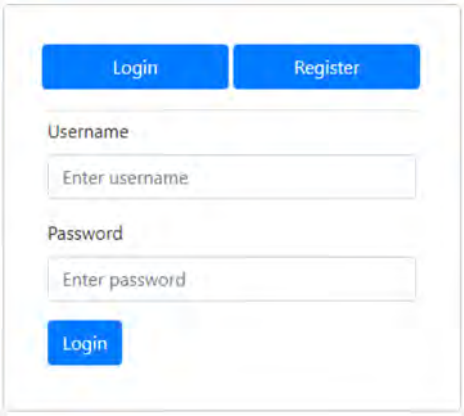


The project works as follows:

- First admin registers all the legit users and saves the credentials of each user in the database.
- When the user wants to enter the system, he has to get authenticated and authorized. For this, he enters credentials.
- These credentials are checked and matched with the ones in the database.
- If the credentials match, the user gets entry in the system else no.
- If the user enters wrong credentials, he is blocked for 5 minutes so that brute force attack can be avoided.
- After logging in a unique ID is generated known as socket ID and it is assigned to each user for concurrency control.
- User sends messages by selecting a user from the chat list and these messages in the database and all of them are sent in first in first out way.
- Before sending the message it will check if the message contains any confidential terms, if yes then the message won't be sent and instead admin will get notified. If no then the message will be sent.



Following are the screenshots of UI which is kept very simple and lightweight to increase the speed so that it can be even accessed by 2G network. Register option will only be available to the admin. And after user login he will have the access to chat list where he can select the group for communication or individual accordingly.



Following is the schema of the database used in this project which consists of user table which keeps the record of user such as username, password, socket-id, online time and other table defines the route of the message and the message. In future scope we can apply AI and deep learning to identify spam messages or confidential messages which are not supposed to be communicated.

1. user Table structure :

Field	Type	Null	Key	Extra
id	int(11)	NO	PRI	auto_increment
username	varchar(10)	NO		
password	varchar(20)	NO		
online	enum('N','Y')	NO		
socketid	varchar(20)	YES		

2. message Table structure :

Field	Type	Null	Key	Extra
id	int(11)	NO	PRI	auto_increment
from_user_id	varchar(45)	YES		
to_user_id	varchar(45)	YES		
message	text	YES		

The following snapshots represent the visuals of the database used in this project using WAMP Server and MySQL database.

	id	from_user_id	to_user_id	message
Edit Copy Delete	1	2	1	Hey
Edit Copy Delete	2	3	1	Hey
Edit Copy Delete	3	1	3	Hi Bro
Edit Copy Delete	4	2	1	It works !
Edit Copy Delete	5	1	2	Yeah it does :)

Check all With selected: Edit Copy Delete Export

After completion of this project we performed penetration testing on this so that we can secure it from the well-known OWASP top 10 vulnerabilities that are

1. Lack of Validation
2. Insecure Configuration Management
3. Broken Access Control
4. Broken Authentication and Session Management
5. Cross Site Scripting
6. Buffer Overflow
7. Injection Flaws
8. Improper Error Handling
9. Insecure Storage
10. Application Denial of Service

Also we have done necessary steps to prevent it such as blocking the user to access the platform if wrong credentials are entered multiple times also input value is escaped and filtered to prevent XSS and other attacks.

V. RESULT AND DISCUSSION

We successfully achieved to eliminate the disadvantages of other messaging systems such as advertisements, distraction because of friends or other spam messages, fake rumors.

We managed to eliminate the subscription system and instead of paying in repetitive manner client can own this with one time pay and can get customized according to its needs which is not provided in others.

VI. CONCLUSION

Thus by this we conclude that the proposed system (Group Messaging Solution) transparency will be there in a hierarchy. Also not only an industry but also other organizations where hierarch y is available. This system helps the group members to have a better communication amongst them and so that everyone can keep their idea and eliminate the research gap.

We first went for literature survey in which we understood several research gaps. There are several systems used for messaging but they have flaws. We conducted surveys to understand the advantages and disadvantages of those applications that the users face. Advantages and disadvantages were analyzed and feasibility study was made to understand what we can do to get better application.

So we tried and created application that is customizable as per what the user needs. The user can anytime think of changes and change the system. The User interface helps the user understand about the working of the system. Our work was based on agile methodology so we completed works in parts and then collaborated them. It removes annoying issues like advertisements. It also has news feed and message detection system which will help the employees stay on the same page and eliminate the spam and rumors and also admin can keep an eye on confidential message. The application can be used all time after paying once.

We successfully matched what was expected and we what is created.

Future Scope:

Through our project we have tried to fulfill all the technological gaps that we found out through the literature surveys and surveys that we conducted. But the project has future scope because it is an interactive project and can be advanced as per the needs and demands of the stakeholders. The following features can further be added to the project:

- Identify messages with priority.
- OCR technology to scan document and send it in text format.
- Adding dictionary.

VII. ACKNOWLEDGEMENT

We sincerely thank to our guide Dr. Zahir Aalam for his guidance and support for carrying out our project work. It is indeed a matter of pleasure and privilege to be able to present this project on **Group Messaging Solution**.

The completion of the project is a milestone in student's life and its execution is inevitably in hands of guide. We are highly indebted and appreciation of the project guide **Dr. Zahir Aalam** for his invaluable guidance and appreciation for giving form and substance to this report. It is due to his enduring efforts, patience and enthusiasm, which has given a sense of direction and purposefulness to this project and ultimately made it a success.

We would like to express our deep gratitude to the staff members for co-operation. Also we would like to thank our Dean R&D **Dr. Kamal Shah** and H.O.D **Dr. Rajesh Bansode** for his sincere support and guidance.

We would like to express deep regards and gratefulness to Principal **Dr. B. K. Mishra** for always encouraging students learn new technologies and implement them.

We would like to thank the non-teaching staff and friends who have helped us all the time in one way or the other.

Really it is impossible to repay the debt of all the people who have directly or indirectly helped us in completing the project.

References

[1] Noor Sabah, Jamal M. Kadhim and Ban N. Dhannoon, "Developing an End-to-End Secure Chat Application," IJCSNS vol. 17, no. 11, November 2017.

[2] Manoj Kumar Srivastava, "Systems and methods for detecting and/or handling targeted attacks in the email channel," US Patent No:US 9,686,308 B1, Date of patent :June 2017.

[3] Joon S . Park, "Role - based access control to computing resources in an inter - organizational community," US Patent No:US 9 ,769,177 B2, Date of patent :September 2017.

[4] Zhen Wang, Zheofeng Ma, Shoushan Luo, Hongmin Gao, "Enhanced Instant Message Security and Privacy Protection Scheme for Mobile Social Network System," IEEE, March 2018.

[5] Deeksha K,P.Hemashree " Intranet Chatting System," IJETER, vol. 6, no. 4, April 2018.

E-KANBAN SYSTEM FOR SUPPLY CHAIN

Arisht Bhandari, Anand Jha, Tushar Jha
Thakur College of Engineering and Technology, Mumbai,India
arisht.bhandari.ab@gmail.com, jhaanand9716@gmail.com, jhatushar123@gmail.com

Abstract— **For years, many manufacturers had faced problems such as manual entry errors and lost cards. Therefore, they have implemented E-kanban system. It can be integrated into ERP (Enterprise Resource Planning) system, which enable real time- demand signal to supply chain and it will improve visibility. Data pulled from E-Kanban systems can be used to optimize inventory levels by better tracking supplier lead and replenishment times. Unlike traditional Kanban which uses traditional elements such as Kanban cards with barcodes and Electronic messages such as electronic data enterchange or email , Electronic Kanban is a signaling system which uses mix technology to trigger the movement of materials within a manufacturing facility. There are two most important types of Kanban system: production Kanban and transportation Kanban.**

I. INTRODUCTION

For years, many manufacturers had faced problems such as manual entry errors and lost cards. Therefore they have implemented E-kanban system. It can be integrated into ERP (Enterprise Resource Planning) system , which enable real time- demand signal to supply chain and it will improve visibility.
Data pulled from E-Kanban systems can be used to optimize inventory levels by better tracking supplier lead and replenishment times.

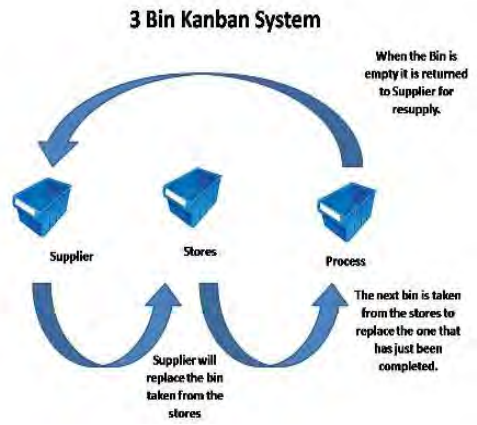
Kanban became an effective tool to support running a production system as a whole, and an excellent way to promote improvement. Problem areas are highlighted by measuring lead time and cycle time of the full process and process steps. One of the main benefits of kanban is to establish an upper limit to work in process inventory to avoid overcapacity. Other systems with similar effect are for example.[5] A systematic study of various configurations of kanban systems, of which CONWIP is an important special case, can be found in Tayur (1993), among other papers.

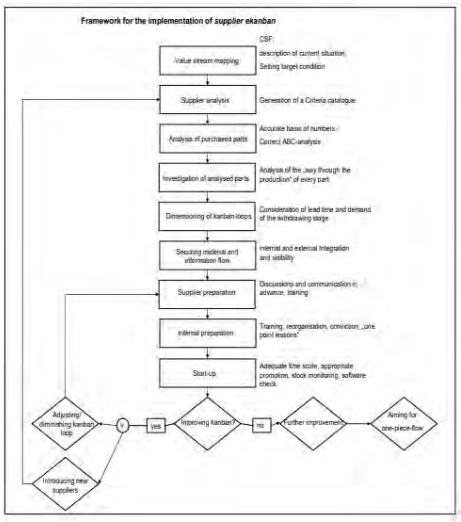
A goal of the kanban system is to limit the buildup of excess inventory at any point in production. Limits on the number of items waiting at supply points are established and then reduced as inefficiencies are identified and removed. Whenever a limit is exceeded, this points to an inefficiency that should be address

II. PROPOSED WORK

In this project we are going to create an application where the supplier falls short of his product, he sends the request through the application to the manufacturer. Then the system will register the requested product. Then manufacturer processes the requested product, and then sends a signal to the supplier regarding the completion of the product. It notifies the supplier to collect the product

The working is as follows:





III. LITERATURE SURVEY

In this paper [1], The kanban contains information that serves as a work order.It gives information concerning what to produce, when to produce, in what quantity, by what means and how to transport it.

In this paper[2] Current production literature alludes to the idea that, collectively, these advances have paved the way for application of Just-In-Time (JIT) production concepts, which were originally developed for mass production systems, in intermittent production systems.

This paper[3] This research work focuses on study of supply chain management of leading Indian and Multinational companies in lighting industry in India in order to understand how they are coping up with the forces of continuity and change.

IV. FEASIBILITY STUDY

- **Economic Feasibility (EFS):** The cost and benefits related with the implemented project compared and the project is economically feasible or not and it benefits outweigh costs. The system development cost will be significant. So the project is economically feasible.
- **Business Case**

  The Business Case provides an analysis of the business environment including-
  Expected customers: supplier, manufacturer.

I. Technical Feasibility:

Resources Required:
Software required: Android studio, angular js.
Hardware required: PC for development, smartphone for implementation.

V. Results and discussion

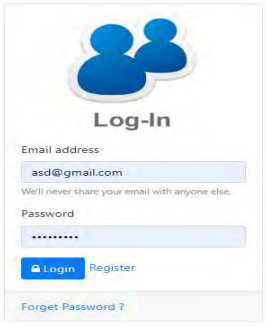


Fig 2 Manufacturer Login Page

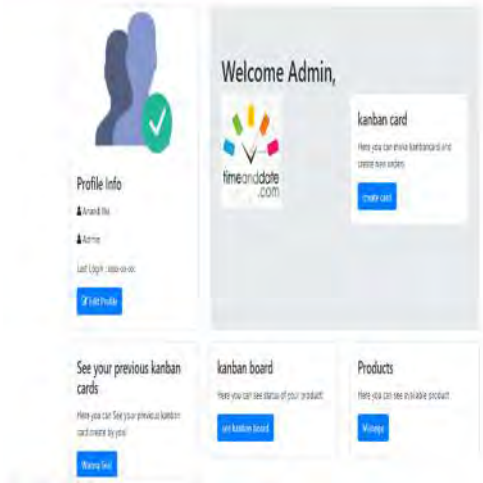


Fig 3 Manufacturer's GUI

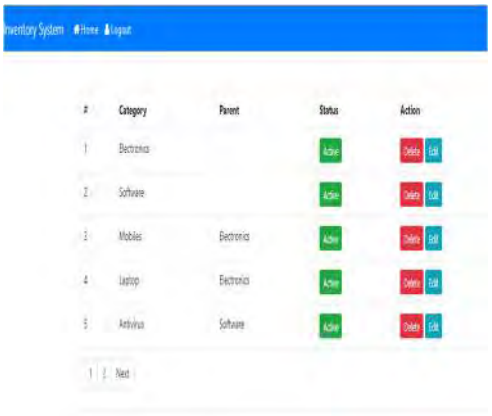


Fig 4. Product Categories

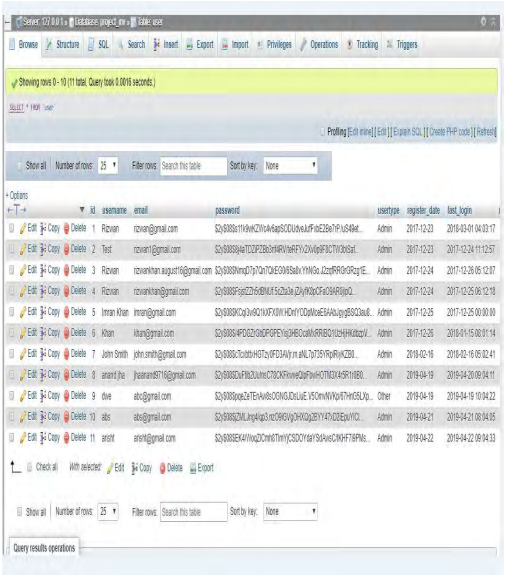


Fig 5. Admin Database

Fig 6 KANBAN Card Details

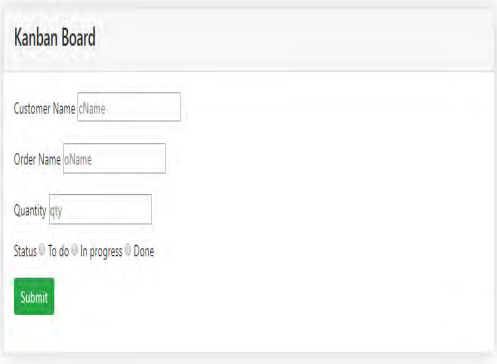


Fig 7. KANBAN Board

VI. CONCLUSION AND FUTURE SCOPE

Framework developed for the implementation of a supplier e-Kanban enhances operational tasks such as providing materials and guaranteeing a smooth run. The framework also allows strategic aims to be followed, and allows any underlying problems that may impact success to come to the surface, to be solved.

We had thus developed an internet-based application as a base for supplier e-Kanban. The application is password-secured and only for suppliers applying an e-Kanban system. By logging in to the system, every supplier can monitor the Kanbans, the different status of the Kanbans, and the current location of the container. Thus, the supplier has an overview of what will be needed shortly and can therefore prepare material in advance, before the actual signal arrives. This clarity gives the supplier more flexibility to ready the next Kanban.

With reference to reviewed work we were able to find and define the problem definition we were able to find and define problem definition. We were clear with the idea of problem definition, algorithms, tools (front end, back end) required to implement the same in phase I of the project.

We have thus achieved our goal and developed the desired system for the implementation of a supplier e-Kanban enhances operational tasks such as providing materials and guaranteeing a smooth run. The framework also allows strategic aims to be followed, and allows any underlying problems that may impact success to come to the surface, to be solved.

References

[1] Ramzan Talib,"*Text Mining: Techniques, Applications and Issues"* (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7 No. 11, 2016.

[2] Ning Zhong, Yuefeng Li, and Sheng-Tang Wu,"*Effective Pattern Discovery in Text Mining*" , Ieee Transactions On Knowledge And Data Engineering, Vol. 24, No. 1, JANUARY 2016.

[3] Abhishek Kaushik,"*A Comprehensive Study of Text Mining Approach*" in IJCSNS International Journal of Computer Science and Network Security, VOL.16 No.2, February 2016 .

[4] Bhavna Bhardwaj, "*Text Mining, its Utilities, Challenges and Clustering Techniques*" ,International Journal of Computer Applications (0975 – 8887) Volume 135 – No.7, February 2016

[5] Un Yong Nahm, "Text Mining with Information Extraction"at 4th International MIDP Colloquium,2013.

[6] Apreutesei, M., Arvinte, IR., Suciu, E., and Munteaunu, D., 2010. Application of Kanban System for Managing Inventory. *Bulletin of the Transilvania University of Brasov,* 3, 52..

[7] Ardalan, A and Diaz, R., 2011. NERJIT: Using Net Requirement Data in Kanban-Controlled Jumbled-Flow Shops. *Production and Operations Management,* 21(3), 606-618.

[8] Barkmeyer, E.J., 2007. An Ontology for the e-kanban Business Process. *National Institute of Standards and Technology,* Internal Report 7404.

[9] Bendavid, Y. Boeck, H. and Philippe, R., 2010. Redesigning the replenishment process of medical supplies in hospitals with RFID. Business Process Management Journal, 16(6), 991-1013.

[10] Cagliano, R., Caniato, F. and Spina, G., 2006. The linkage between supply chain integration and manufacturing improvement programmes. *International Journal of Operations and Production Management,* 26 (3), 282-299.

[11] Chen, K-M. Chen, J.C. and Cox, R.A., 2012. Real time facility performance monitoring system using RFID technology. Assembly Automation, 32(2), 185-196.

[12] Chong, A.Y.L, Ooi, K.B, Murali. R. and Lin, B., 2010. Factors affecting the adoption level of collaboration commerce technologies in supply chain management: an empirical study of Malaysian electrical and electronic organizations. *Journal of Computer Information Systems*, 50(2), 13–22.

[13] Christopher, M., 2011. *Logistics and Supply Chain Management*, *y*4th edition, Prentice Hall: London.

[14]Cimorelli, S., 2005. *Kanban for the supply chain: fundamental practices for manufacturing management*. 1st Edition, Productivity Press: US.

[15] Dickmann, P., 2007. *Schlanker Materialfluss*., 1st Edition, Springer- Verlag Berlin: Heidelberg.

[16] Drickhammer, D., 2005. The Kanban E-volution. *Material Handling Management,* March 2005, 24-26.

[17] Eisenhardt, K.M. 1989. Building theories from Case Study Research. *The Academy of Management Review*, 14(4), 532-550.

Student Attendance System using QR Scan Code

Tanvi Haldipur, Adarsh Kotian, Sumer Patel
Information Technology, Thakur college of Engineering and Technology,Kandivali
Mumbai, India

Abstract – **Most of the teachers/professor's waste about most of their time of the lecture taking the attendance for their students. Student attendance using QR scan is an application that is developed in order to ease out the life of teachers. This app will be able to take the attendance of the students with the help of QR codes. Nowadays most of the students have their own smartphone or at least a phone that could open web browsers, so availability of smartphones is not a major problem here.**

The day to day register work of the teachers is reduced. The teachers tend to spend most of their time taking the attendance using the old process of manually entering data into the registers. With the help of this application teachers will be able to enter the details of the student, the time the students QR code was scanned, the student ID all these data will be stored into the sheets.

This app will be a web app as well as it could be opened in any browser of your choice. The students will generate their QR codes by entering their data (the data will consist of the student's unique identification number, students full name and the time the students QR was scanned), which are generated dynamically so as to prevent any proxy. The teacher only will have a scanner application who will scan the dynamically generated QR codes and the data that is entered by the student is stored in the data sheet of the teacher for attendance purpose.

I. INTRODUCTION

Attendance Management System is software developed for daily student attendance in schools, colleges and institutes. It facilitates to access the attendance information of a particular student in a particular class on daily basis. The information is sorted by the application, which will be provided by the teacher for a particular class or classes. This system will also help in evaluating attendance eligibility criteria of a student for any remedial work need to give for monthly attendance generated by the report generated by the attendance application system. Problem definition:

This system will generate attendance of using QR Scan and generate reports to professor regarding monthly attendance and give a detailed report of all the defaulters in particular lecture in particular dates the students remained absent. This is hassle free system in which no use of pen and paper is required and saves time as mobiles are used by each and every body in today's world so once this application is installed it is convenient for professor as well as for student and even no proxy attendance will be marked.

Background:

Traditionally, the student attendance was taken in a manual way which was either using the formal or informal paper attendance sheet. The manually taken student attendance is uncomfortable in teaching and

II. METHODOLOGY USED

Learning situations to lecturers even to students. This project is developed to manage and reduce the current problems faced by the lecturers and students in a systematic and structured method. The system only can be functional after it is activated by the administrator or lecturer to identify and record students' attendance. This can prevent the student or other people to access it. There are some functions provided including adding and editing student attendance information, calculating the percentage of student attendance and printing the student attendance report adding leaves regarding medical if absentee. period and tasks involved with them. To remain in sync with the other members of the team they have to complete their task within the stipulated amount of time.

The meetings however are conducted on a regular basis where all the members of the team have the right to put forth their ideas along with any requirement or concern. The product backlogs are discussed in these meetings where the requirements of the project to be completed along with the resources are mentioned.

The project is first divided into different phases, and then the responsibility of developing these phases is given in the hands of the various developers. At every stage of the project development it is continuously monitored and taken care of. These developers developing the phases have particular amount of time. Agile

Agile methodology mainly focuses on the incremental and the iterative development of the different project phases. It controls and determines how the various

III. TECHNICAL WORK/RESEARCH

Feasibility Study

1. Executive Summary

Project is related to Student Attendance System using QR Scan.

The project maintains two levels of users: - Administrator Level-Owner

User Level-Data Entry Operator

Main facilities available in this project are: - Maintaining records of student's attendance. Maintaining monthly attendance details. Providing different professor the attendance.

Maintaining backup of data as per user requirements (between mentioned dates).

User or Administrator can search a student record by his/her name

2. Define business problem or opportunity

The Software is for the automation of Attendance Management.

It maintains two levels of users: - Administrator Level

User Level
The Software includes: - Maintaining student details.
Providing and maintaining all kinds related to attendance for professor ease.
Auditing and Report generation.

3. Requirements and purpose of the study
Hardware Requirements Processor RAM, Disk Space Pentium IV or higher Software Requirements Operating System.
Win-XP or higher.
MS Excel.
Web Development tools.
External APIs to generate QR codes dynamically. Purpose QR Scan provides facilities like: - Taking attendance easily.
Saves time
Saves pen paper usage
Record maintenance easily done

4. Description of the options assessed
Based on research conducted over the Internet we identified the following alternatives for the MMS:

Continue to use the current system. This will be referred as Alt 1.
Utilize a packaged solution. The colleges and schools who require this software can purchase the product and host them in their own environment. A lot of programming and re-programming is required for customizing the system and implementing them. Since this is a proprietary system the programming costs are at a premium. This will be referred to as Alt 2.
Develop a customized MMS solution which meets all requirements satisfactorily and is also built with a scalable scope. The programming costs will be on a regular cost. This will be referred to as Alt 3.

5. Assumptions used in the study
Every user should be:
• Comfortable of working with computer.
• He must have knowledge in developing Application for QR Codes.
• He must also have basic knowledge of web and development skills.

6. Audience impacted
It can be used in any attendance management from colleges, schools or workplace.

7. Financial obligation
The financial and the economic questions during the preliminary investigation are verified to estimate the following:
The cost to conduct a full system investigation.
The cost of hardware and software for the class of application being considered.
The benefits in the form of reduced cost.
The proposed system will give the minute information, as a result the performance is improved which in turn may be expected to provide increased profits.

8. Recommended action
Time evaluation is the most important consideration in the development of project.
The time schedule required for the developed of this project is very important since more development time effect machine time, cost and cause delay in the development of other systems.
A reliable Student Attendance System using QR Scan can be developed in the considerable amount of time

IV. DESIGN AND IMPLEMENTATION

4.1 DFD:
Level 0:

Fig: 4.1.1: DFD Level 0

Level 1

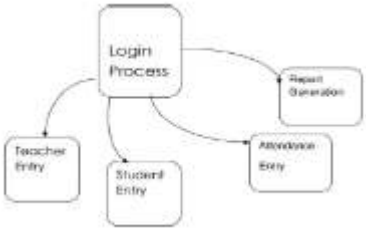


Fig: 4.1.2: DFD Level 1

Level 2:

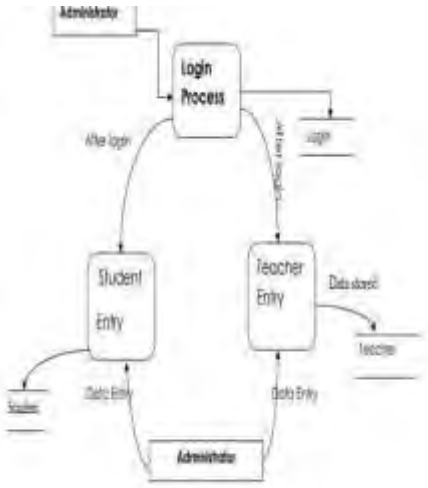
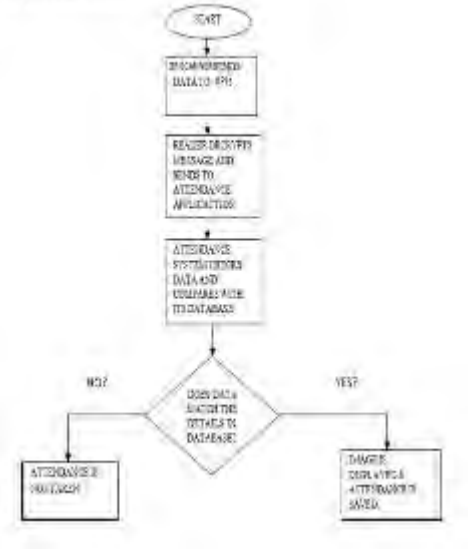


Fig: 4.1.3: DFD Level 2

4.2 Flow chart:



V. RESULT AND DISCUSSION

After the complete analysis done through the previous sections of the paper. The results related are displayed below.

Actual output:

As far as Attendance is concerned a feedback system can be included. A section where students can download their assignments right from the website or WebApp also Push Notifications/SMS/Email of any changes that occur in the day to day schedule.

VI. CONCLUSION

With reference to reviewed work we were able to find and define the problem definition according to the methodology implemented.

We are clear with the idea of problem definition, algorithms, tools (front end, back end) required to implement the same.

We will follow phases shown above to implement proposed project for smooth development conduct.

We will try our level best to implement the proposed system to achieve the defined objectives.

VII. FUTURE SCOPE

As far as Attendance is concerned a feedback system can be included.

A section where students can download their assignments right from the website or WebApp also Push Notifications/SMS/Email of any changes that occur in the day to day schedule.

Style of list in references of some standards is as below;

References

[1] Text Book References

- Web Development with Bootstrap 4 and Angular 2 by Packt Publishing.
- Scan Me - Everybody's Guide to the Magical World of QR Codes by Mick Winter.

[2] Journal references
IEEE standard references format

[1] IJACSA, 2014, Volume 5, Issue 3, April 2014, Fadi Masalha, Nael Hizallah

[2] IJIIT 2016, Volume 5, Issue4 April 2016- 2017, Seema Padalkar, Smita Sawakare, Ankita Salwade, Shradha Wankhade, Chhaya Pawar

[3]AJAST 2017, Volume 1, Issue 5 June 2017, B.Dinesh Kumar and S.Kareemulla

[3] Web references

- https://www.researchgate.net/figure/Flow- Chart-of-the-QR-Attendance-Monitoring-System_fig5_315706580
- https://meeraacademy.com/use-case- diagram- student-attendance-management- system/

Automated Healthcare Informatics On Service Management

Mansi Agarwal, Irshad Badarpura, Bhoomi Gandhi,Vikas Kaul
Information Technology, Thakur College of Engineering and Technology/ Mumbai University, India

Abstract – **Health care is the essential part of our lives, providing best medical facilities to people suffering from various ailments, which may be due to change in climatic conditions, increased work-load, emotional trauma stress etc. It is necessary for the Health care industry to keep track of its daily activities & History of its patients, doctors, nurses, ward boys and other staff personals that keep the Health care running smoothly & successfully. But keeping track of all the activities and their history on paper is very strenuous and error prone. It also is very inefficient and a time-consuming process. Observing the continuous increase in population and number of people visiting the Health care. Recording and maintaining all these records is highly unreliable, inefficient and error-prone. It is also not economically & technically feasible to maintain these records on paper. Thus, keeping the working of the manual system as the basis of our project.**
We are using a tableau tool for developing dashboards which help to eliminate the need of paper based visualization and help the Health care business users to effectively analyze and monitor the daily takes that are performed in the Health care. Basically, we use the data mining and business intelligence techniques and methods to provide the stakeholders with detailed inspection of the Health care workings. Also, we use a tool called Pentaho that provides business intelligence and OLAP services, reporting, Data mining and ETL Capabilities. Using this we extract only the data needed for analysis from the huge data sets.
In the future, this project can be extended to give more detailed analysis of the health care industry and solve problems such as maintaining manual record of data, studying the data manually and can help to save a lot to time. Also, it can be used to predict the outcomes and suggest measures that can be taken to prevent an error and significantly it will help to reduce labor costs.

I. INTRODUCTION

Healthcare are the essential part of our lives, providing best medical facilities to people suffering from various ailments, which may be due to change in climatic conditions, increased work-load, emotional trauma stress etc. It is necessary for
the Healthcares to keep track of its day-to-day activities & records of its patients, doctors, nurses, ward boys and other staff personals that keep the Healthcare running smoothly & successfully. But keeping track of all the activities and their records on paper is very cumbersome and error prone. It also is very inefficient and a time-consuming process Observing the continuous increase in population and number of people visiting the Healthcare. Recording and maintaining all these records is highly unreliable, inefficient and error-prone. It is also not economically & technically feasible to maintain these records on paper. Thus keeping the working of the manual system as the basis of our project. We are using a tableau tool for developing dashboards which help to eliminate the need of paper based visualization and help the Healthcare
business users to effectively analyze and monitor the daily takes that are performed in the Healthcare. Basically we use the data mining and business intelligence techniques and methods to provide the stakeholders with detailed inspection of the Healthcare workings.

II. II. Material And Methods

This prospective comparative study was carried out on patients of Department of general Medicine at Dr. Ram Manohar Lohia Combined Hospital, Vibhuti Khand, Gomti Nagar, Lucknow, Uttar Pradesh from November 2014 to November 2015. A total 300 adult subjects (both male and females) of aged ≥ 18, years were for in this study. (10) Study Design: Prospective open label observational study
Study Location: The study was done at IT Department, Bhakti Vedanta Hospital, Mira road,Thane,Maharashtra.
Study Duration: July 2018 to November 2018.
Sample size: 1 million patients.
To create the dashboards we extract the data from different data sources and data marts using a tool called Tableau and various Data warehousing techniques.
The different methodologies used are :

E-R Model:

It is a graphical illustration of entities and relationship to each other, usually used in computing in reference to the organization of data within databases or information system.

ETL process:

It is used to extract the data from different relational database sources then transform the data using calculations, concatenations, etc. and then load into the Data warehouse system.
The various technologies used are:

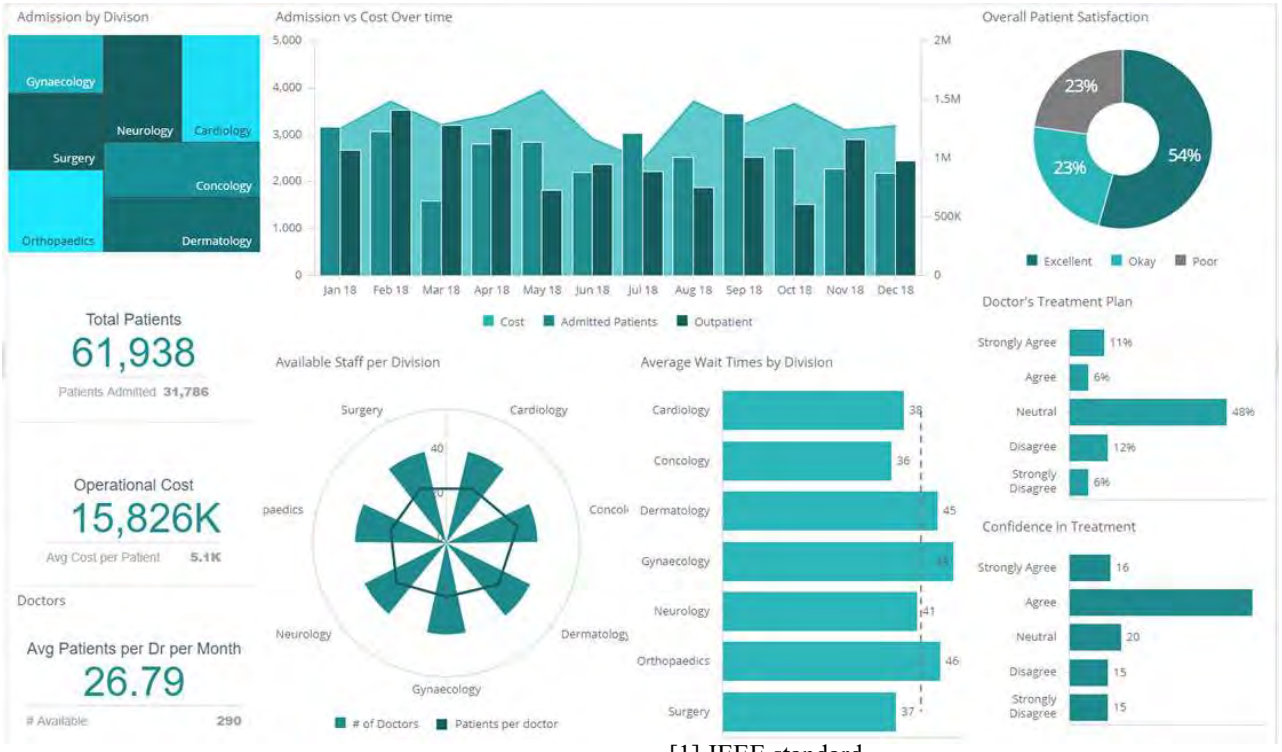
Tableau:

Tableau is a tool that helps to understand and analyze data. Users can generate and dispense an interactive and
shareable dashboard, which illustrate the trends, variations, and density of the data in the form of graphs and charts. Tableau can bind to files, relational and Big Data sources to acquire and process data. The software allows data blending and real-time combination, which makes it very eccentric. It is used by businesses, academic researchers, and many government organizations for visual data survey.

Pentoha:

Pentoha is a tool for creating relational and analytical report. It helps to create reports in HTML, Excel, PDF,
Text, CSV, and xml. Pentaho can take data from variety of data sources including SQL databases, OLAP data sources, and even the Pentaho Data Integration ETL tool.

III. III. Result



IV. IV. Conclusion

With reference to reviewed work we were able to find and define the problem definition we were able to find and define problem definition. We are clear with the idea of problem definition, algorithms, tools (front end, back end) required to implement the same. We will follow phases shown above to implement proposed project for smooth development conduct. We will try our level best to implement the proposed system to achieve the defined objectives.

[1] IEEE standard

References

[1] Data Analytics in Healthcare Informatics , Fei Wang ; Gregor Stiglic ,2015 International Conference on Healthcare Informatics
[2] In Technology We Trust: Extending TAM from a Healthcare Technology Perspective
[3] Rupali Jain, Clinic Automation System: An Example of Medical Informatics , International Journal of Computer Applications (0975 –8887) Volume 137 – No.12, March 2016
[4] Raja Manzar Abbas ; Noel Carroll ; Ita Richardson 2018 IEEE International Conference on Healthcare Informatics (ICHI)

Intelligent Tourist System For Better Travel Experience

Sumanth Bhat, Nitin Ojha, Aditya Pandey, Neha Patwari
Department of IT Engineering Thakur College of Engineering amd Technology
Mumbai,India
sumanthbhat112233@gmail.com, nitinojha613@gmail.com, pandeyaditya.it@gmail.com, pandeyaditya.it@gmail.com

Abstract— **Nowadays people use mobile phones and other mobile devices. Most of us have a small computing device that is always with us. People use it example for calling, as calendar and organizer. Mobile devices with GPS receiver are also used to find paths in navigation.**
The main idea of this thesis was to design a system that will run on most of phones and palms and will be helpful when visiting some new places and cities. This system should be able to find a route using user criteria.
The project belongs to the domain of Machine Learning and
Artificial Intelligence and the programming language used is python.
Many papers have been published, research have been conducted to analyze the methods to be used for developing a better intelligent tourist system and based on some of
those paper this project includes various points
Intelligent tourist system for better travel experience will enable the travelers to carry out their journey in an efficient manner. This project covers all the essential factors needed for traveling.
The additional functionality the project includes is videos from famous bloggers related to exploring new places. This functionality will help travelers to explore new places with ease. By using this feature the travelers can take some knowledge about the concerned places before visiting them.

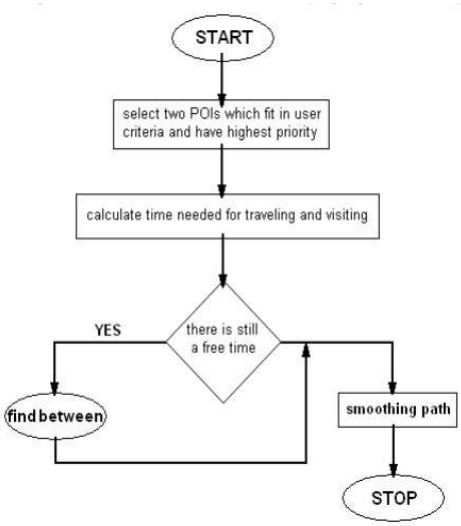
I. INTRODUCTION

The project aims at building an intelligent tourist system to
assist the travelers in exploring various tourist places. In the past number of applications have been developed for same. The Primary reason was to find an efficient manner of traveling. With past studies and researches it has been of major concern to develop a system which could help travelers while traveling. A lot of papers have been published with real time model having intelligent tourist system embedded in them.
accurate measurements but their cost is usually as high as several hundred dollars, making them uneconomical.
So this heart rate monitor with an optical sensor is a very useful instrument in knowing the pulse the patient..

II. PROPOSED WORK

The methodology that we chose for our project is Agile. While waterfall model is a tried and tested framework, it is not suitable for our project, due to the changeable nature and small team size. Agile allows us to be flexible, manage changing requirements, manage the ever-increasing scope as well as get consumer perspective.

The working is as follows:



III. LITERATURE SURVEY

[1]Literature Survey on Tourism Recommendation
Jitha P B, Reshma M
Recommender system, comparative description, mobile system, topic modeling
[2]Intelligent Tourist Information System
AyeshaShaikh,Sonal Rami,
Tirth Shah Recommendation technique, Shortest pathalgorithm, Intelligent tourist system, A* algorithm.IV. FEASIBILITY STUDY

- **Economic Feasibility (EFS):** The cost and benefits related with the implemented project compared and the project is economically feasible or not and it benefits outweigh costs. The system development cost will be significant. So the project is economically feasible.
- **Business Case**

The Business Case provides an analysis of the business environment including-
Expected customers:
The nature of the business:

IV. Technical Feasibility:

Resources Required: Programmers, Arduino.
Software required: Thing Speak, Arduino IDE
Hardware required: PC for development, hardware components for implementation.

V. Results and discussion

Thus after the implementation of prototype we finally come to a conclusion and discussion that the cost of our project or system has to be reduced and efficiency has to be improved as well. The main change to focus on is the size of the system. It has to be compact and easy to use in daily lives of users. So in next phase our primary goal is to focus on size and efficiency of the proposed system.

VI. CONCLUSION AND FUTURE SCOPE

Thus by this we conclude that the proposed system can be efficiently used by travelers and any normal person. This System provide all the traveling related information.

This will provide a way for efficient solutions for travelers economically in terms of money and time.This project also provides a knowledge of

References

[1]. Literature Survey on Tourism Recommendation Jitha P B, Reshma M P.G Scholar, Dept. of Computer Science and Engineering, Cochin College of Engineering, Kerala, India Assistant Professor, Dept. of Computer Science and Engineering, Cochin College of Engineering, Kerala, India

[2]. Research Design of Intelligent Tourist Guide System and Development of APP Kaijian Huang1, a and Junwu Zhu2,b School of Information Engineering, Yangzhou University, Yangzhou, China 2 Department of Computer Science and Engineering, Nanjing University, Nanjing, China

[3]. International Journal of Computer Applications (0975 –8887) Volume 175 – No.3, October 2017 6 Intelligent Tourist Information System Tirth Shah B. Tech. Dept. of IT, Charusat Sonal Rami Assistant Professor Dept. of IT, Charusat Ayesha Shaikh Assistant Professor Dept. of IT, Charusat

[4]. DESTINATION INFORMATION MANAGEMENT SYSTEM FOR TOURIST Abdulhamid Shafii Muhammad1, Gana Usman2 Dept. of Cyber Security Science, Fed. Univ. of Techno. Minna, Niger State, Nigeria.1 Department of Physics, Bayero University Kano, Kano State, Nigeria.

A Review on Comparative Analysis of Machine Learning and Deep Learning Techniques.

Vikas Tiwari Bhushan Nemade
Department of Information Technology
Thakur college of Engineering and Technology
Mumbai-400 101
v8kgsm@gmail.com, bnemade@gmail.com

Abstract. **Now a day's in Industries a hot topic as a trend has taken place termed as Data Processing Techniques Like Artificial Intelligence (AI), Deep Learning (DL) and Machine Learning (ML)[2]. Everyone maybe knowing bit glimpse or maybe they are expert in certain part/technique but one person calls it DL while at the same time the other one calls it ML. This techniques are being used in several sectors by huge com- panies like Google, Facebook, Microsoft etc[3]. As many of the keys and working are same in some perspective of Data Processing techniques but there is key difference between them where we can make differentiation how one Data processing technique is different from other Data processing technique [3]-[5]. In some sectors DL shows the output more exceptionally well were in some sectors ML takes credit of being more perfect. But with respect to computational methodology DL technique is more expensive. In this paper, we have investigated the key difference between Data Pro- cessing techniques and reviewed the comparative analysis of Data Processing tech- niques.**

Keywords. Artificial Intelligence; Machine Learning; Deep Learning; Data Pro- cessing Technique; Comparative Analysis.

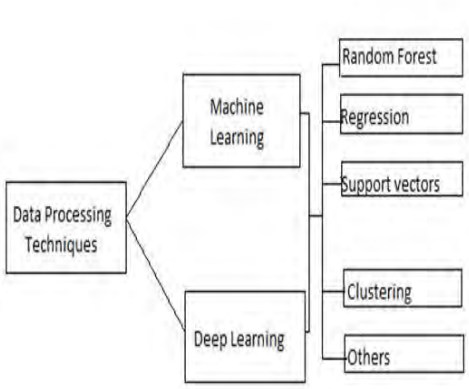
I. INTRODUCTION

Nowadays in many organizations, demand has been boosted due to importance of using predictive analysis or Data Processing techniques. But to know the compatibility of Technique with the System in which we are using it such that it gives the best possible or accurate result we should the key differences between the data processing tech- niques. Lets consider the most general or basic view that what the ML is all about , basically it refers to the computer which can make learning on its own without any manual code implemented on it but an good amount of data is feed to process and make self learning easy. Famous Alan Turing's seminal paper which was proposed in 1950 concludes that Learning Machine can fool a human and make them believe that it is real[1]. About Machine Learning and some part of Deep Learning such as Deep Neural Network (DNN) are playing an important role in making advancement where Machine Learning works on prediction and Deep Learning follows a wise flow of Hierarchy. There are two terms in machine learning i.e. supervised machine learning and unsuper- vised machine learning that is expected to be known by Data Scientists. Ensemble mod- eling uses a combination of approaches and semi-supervised machine learning comb both the techniques. The term supervised machine learning is one in which the user trains the program (Data set is known), such that it act's accordingly as the input is given to it whereas the term unsupervised machine learning generates the answer on unknown dataset. Supervised machine learning commonly uses techniques like Support Vector Machine (SVM), regression, random forests, decision trees, and support vector ma- chines. Unsupervised machine learning uses algorithms like clustering algorithm where Data scientists uses this type of techniques to find or discover new pattern in the given data set. See Fig.1. for basic classification of Data processing Techniques. Deep learn- ing is one of the form of Machine learning where it uses supervised or unsupervised algorithms, or both algorithms has been used. Deep learning is widely used in computer vision, natural language processing, neural networks (It is based on representation learning or feature learning). Google's 'Deep Mind' team has famous for making a con- volutional neural network which beat the ancient

Chinese game known as GO.

.1: Algorithms in Data Processing techniques.

Deep Learning has successfully implemented a strong field ranging from image processing, speech recognition with medical diagnosis to self or auto mate driving cars and as we discussed it is defeating human players in complex games such as Go, chess and etc. As it requires a huge set of data for processing proportionally the requirement of space also increases widely with the complexity of Deep Neural Networks (DNN). For training a DNN we require a high performance computing cluster (categorization of samples based on similarity in feature



II .RELATED THEORY

AntonioHernández-Blanco et al [1] mentioned in his article about Educational Data Mining. Where the terms like Learning Management System, Intelligent Tutoring System, Massive Open online courses. It is concerned with developments, research, and applying ML , data mining and statistical methods to detect pattern in large set of educational data which is used for analyzing the pattern present in the dataset.

Scott H. LeeID et al[2] proposed article on "A comparison of machine learning algorithms for the surveillance of autism spectrum disorder" – The data gathered from site and algorithm has been checked for 10 random train tests where 8 supervised learning algorithms has been applied . It has been noticed that the neural network and random forests almost same result with the accuracy of more than 87 %.

Nitish Srivastava et al[3] proposed Deep neural model concept has a large number of hidden non-linear layers which can solve a very complicated relations and can be implemented with less amount of training data . There are various types of penalties which can come to picture which leads to over fitting of Neural Networks. Dropout Neural Networks can also be trained using stochastic gradient as much similar to standard neural networks.

Yu Wang et al[4] mentioned that the DL based combination of CNNs where it is compatible of recoginising eight modulations with high identification accuracy where it requires a huge amount of data to follow a set of instructon by setting up an hierarchy. The approach can directly be implemented to CRs if more data signals are provided to it. Battista Biggioa,b,∗, Fabio Rolia,b et al[6] Said when a data is fed for the processing the more the experience increases or the learning increases of the machine the more better it will perform with the ability to stop the malware attacks or prevent security threats, most of the times the exact representation layer is not known by the attacker. To prevent the Poisoning attacks the term called data sanitization is used i.e. Detection of attack and simultanoeusly removal of it.

Hadeer Ahmed1(&), Issa Traore1, and Sherif Saad2 et al[7] has done classification techinque of ML and it has been used for Analysis of Fake news called Linear Support Vector Machine with Term Frequency-Inverted Document Frequency (TF-IDF) for feature Extraction. For classification we used ML with less amount of data where the question arises the system is making analysis and showing results with what accuracy, for that they purposed a accuracy results using SVM with different datasets. As the topic was new the good amount and perfect type of datasets were not avilable as they agree to run the model using publicly avilable dataset , Such as LIAR dataset.

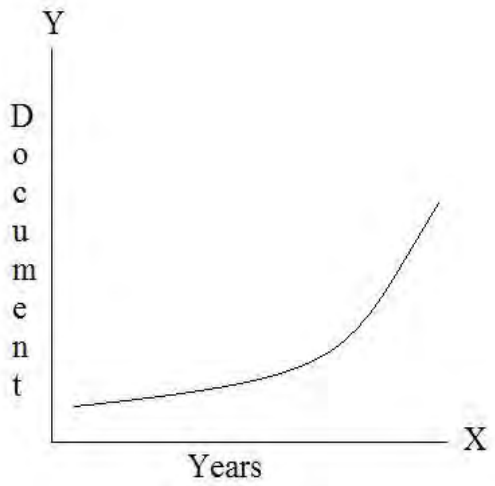
JO˜AO GAMA et al[8] elaborates the terms of Data Processing techniques exists from 90's but the avilability of data was not much therefore the use of such techniques were less but as the time passed , the amount of data keept increasing as the survey conducted according to [8] in 2012 the amount of data reached 2.8 zetabytes. Now a huge amount of data is avilable due to which the predictive and learning adaptivity with accuracy, is boosted. The amount of correct data present is directly proportional to the accuracy of the model. As per Statistics[9] the amount of data in 2020 will reach from 4.4 zetabytes to 44 zetabytes.

CHUANLONG YIN et al[10] mentioned that ML methodologies have been widely used for prevention of various types of attacks where the

technologies help Network Administrator to identify the threats and malware attacks and take preventive measures but it is not successful to solve the massive intrusion problem of data classification which arises in the real network scenario . Where Deep Learning has better potential to identify the representation layer and it achieve a special remarks in speech recognition , image recognition[11] and action recognition[12]-[13] .

Andreas Kamilaris et al[14] introduced Smart Farming with Deep learning, the Deep Learning technique has been used for tackling the challenges of farmming related problems in terms of productivity of the suitable crops with a lot many appropriate conditions. To get addressed and to tackle all the challenges, all the complexity like unpredictable and multivariate ecosystems they required the continuous monitoring support with all aspects which points towards the big data. Which required a imaging analysis with optical and x-ray imaging. With all this it is concluded that DL is the most promising aspect to use rather than that of ML techniques. See Fig.3. to analyse the subparts of techniques. DL extends ML by adding more extension to it.

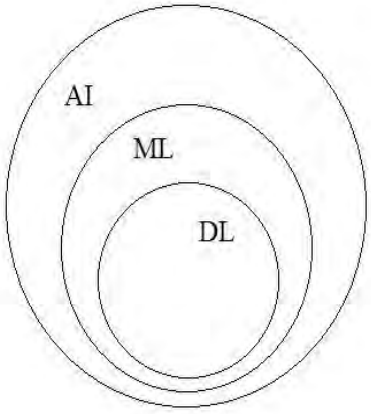
Kamran Kowsari et al[15] said that each year a lot of publications takes place due to which a lot of documentations are generated See Fig.2. , likely to be said that only scientific reasearchers each year produces massive number of documents. In 2014 around 2.4 million papers have been published which



itself is a large amount where every year it keeps on increasing

Fig.2. Graphical representation of incrementation in documentation over the years.

To handle such a large amount of data we need Deep Learning to come into picture where ML cannot handle such large amount of data. It will use



hierarchical document classification where hierarchical DL is used for text classification.

Fig.3. Domain Specifier (what is subpart of what).

II. METHODS

This section ensures that why the analysis is important on Data processing Techniques such that one should know that which technique will enhance the accuracy, compati- bility and endurance of models. With series of analysis we came to know that both the techniques (ML and DL) have their advantages and disadvantages based on the model what one individual want to implement. As such many papers mentioned number of models, which authors are trying to implement or have implemented but the compati- bility depends on the type and amount of data that one individual is going to gain cer- tainty with level of accuracy. In ML the machine is trying to learn with the past expe- rience i.e. it is trying to reduce the difference between the estimated value and the actual value which leads to decrease in the error and try to maximize the accuracy level to give more accurate predictions. For example if we want to predict the height on given data of weight or vice versa. See Fig.4. For reference.

Fig.4. Representation of Predictive analysis.

We can use more than one attribute for prediction for which we use Multi regression technique. The Deep Learning is the particular type of machine learning which uses mostly the same configuration of how our brain functions on neuron level if we want to analyze anything such as box then it follows hierarchical order to analyze the object, it first checks whether it has four sides or not , then it checks all sides are connected or no, further they are perpendicular to each other or no . This is how it analyze all the things on higher level like image processing, speech recognition and much more.

Classifiers	Acc.
RF	**0.9974**
DNN 2L	0.9779
DNN 4L	0.9742
DNN 7L	0.9615
RF	**0.9978**
DNN 2L	0.9884
DNN 4L	0.9869
DNN 7L	0.9620
RF	**0.9941**
DNN 2L	0.9695
DNN 4L	0.9899
DNN 7L	0.9921

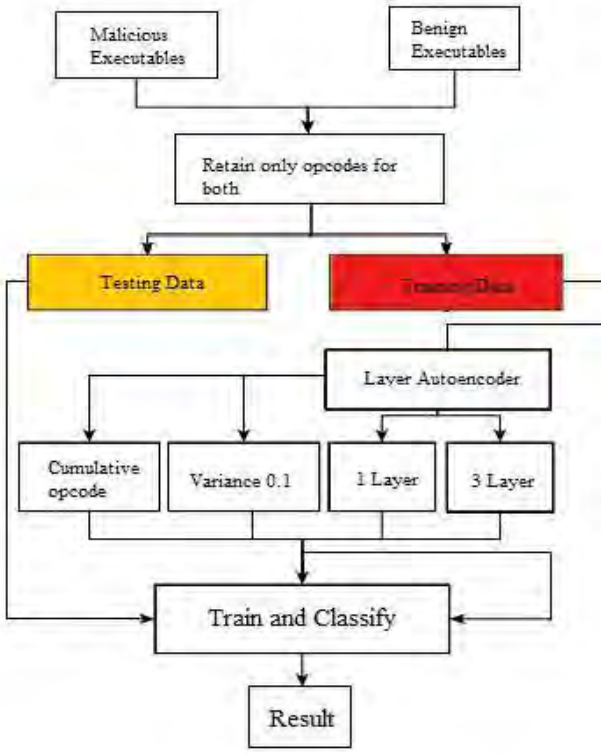


Fig .5. Formula for

IV . COMPARITIVE STUDY

The comparison on the basis of ML classifier such as Random Forest and Deep Learn- ing approaches with enhancement of Deep Neural Network using 2 , 4 and 7 layers. The software used for Random Forest is python library sci-kit[16] and for Deep Neural Network is python library keras[17]. Some layers have been chosen to perform and analysis the performance with increase in the number of layer. The formula for the accuracy See Fig.5.

$$\text{Accuracy} = \frac{TP + TN}{TM + TB}$$

TP -> True positives.
(The number of malware instances correctly classified)
TN -> True negatives.
(The number of benign instanes correctly classified)
TM -> Total number of malware instances.
TB -> True number of benign instances.

Fig.6 Flow chart of comparative analysis

Accuracy.

All DNN layers uses Exponential Linear units to activate the function. To analyze the flow of progress See Fig.6. Where the comparison in the form of flow chart of malware classification by both the techniques i.e. ML and DL. As presented the result for the same analysis the concept of ML classifier i.e. Random Forest has better results than that of 2, 4 and 7 layers of DNN [19]. Although if compared with all the 7 layers of DNN it gives the out-forming result with the accuracy of 99.21% .Refer Tabel.1. to analyze the accuracy of the classifier. If DNN - 4 layers used with AE-1L then it gives the highest accuracy where as if it is used with AE-3L then the accuracy decreases with approximation of 6%. As a result the best accuracy obtained by Random Forest and Deep Neural

Network is 99.78% and 99.21% respectively [18]-[19].

Classification Accuracy for Oasis Cross Sectional (1) and Longitudinal Data (2)

Classifier	Classification Accuracy (1)	Accuracy after Attribute Selection	Classification Accuracy (2)	Accuracy after Attribute Selection
J48	99.52%	99.52%	99.20%	98.66%
Naïve Bayes	99.28%	99.52%	96.78%	98.66%
Random Forest	92.55%	75.96%	90.08%	98.39%
Multilayer Perceptron	96.88%	99.52%	74.53%	97.32%

Table.1. Showing the accuracy of the classifier.

With the reference of table.1. it is clear that Random Forest is more compatible to show high accuracy then that of DNN, but in some cases like a paper proposed by Scott H. LeeID et al[2] on "The Comparison on ML algorithms for the surveillance of autism spectrum disorder" shows that Random Forest and Deep Learning approaches shows the same result .

Due to predictive technique, the Machine Learning has a trend grabs the interest to- wards neurodegenrative disorders. The major emerging global health issue termed as Dementia has been predicted at early stage. Due to which it has become easy to diag- nose Dementia. The predictive analysis with the help of four algorithms named as J48, Naive Bayes,Random Forest and Multilayer perceptron. The raw dataset partly col- lected from oasis-brains.org for processing. The work flow and the comparitive accu- racy factor analysis for each mentioned algorithm is as mentioned in Table.2. for refer- ence.

Classification Accuracy = (Correctly classified samples / Total number of samples) * (100)

ALGORITHM	ARTIFICIAL NEURAL NETWORK	SUPPORT VECTOR MACHINE	DECISION TREE
Big data based	Yes	Yes	Yes
Water quality factors	physical,bio,chemical	chemical	Biological , chemical
Structured data sets	Yes	Yes	Yes
Training data	60%	78%	70%
Testing data	20%	45%	30%
Real-time prediction	No	Yes	Yes
Simplicity	Yes	Yes	No
Accuracy	87%	Err 0.089	85%
Regression based	Yes	No	No
Supervised Learning	Yes	No	Yes
Sensors Used	No	Yes	No
Robustness	Yes	Yes	No
Flexibility	Yes	No	Yes
Data clustering	possible	possible	impossible

Table.2. Classification Accuracy for types of data.

The updated result for the comparison of ML algorithms based on the factors such as Big-data, water quality, structured datasets, training data, testing data, Real time pre- diction, simplicity, accuracy, Regression and many more. See Table.3. for analytics. To compare the algorithm's efficiency.

Table.3. Comparison of Algorithms Efficiency of ML Algorithm's of Machine Learning namely Artificial neural network, Decision tree, Sup- port vector machine.

V. CONCLUSION

If we want to select which network topology is best like bus , star , mesh or other we can't chose it because each has their pros and cons like wise if we want to chose be- tween AI , ML or say DL, we can't be sure for particular as mentioned they have their pros and cons. Further in future it is only going to keep increasing as the amount of data going to increase in future and the accuracy of the predictions and power of analysis will boost and a great artifact future is waiting for us as likely it is said that further more than oxygen , Artificial intelligence will become our need for survival. We will analyze and make more comparison when the accuracy will boost , with the help of models as per available technique

References

1. AntonioHernández-Blanco ,1 BorisHerrera-Flores ,2 DavidTomás ,3 andBorjaNavarro-Col- orado," A Systematic Review of Deep Learning Approaches to Educational Data Mining," received 28 december 2018- https://doi.org/10.1155/2019/1306039
2. Scott H. LeeID*, Matthew J. Maenner, Charles M. Heilig," A comparison of machine learn- ing algorithms for the surveillance of autism spectrum disorder," online - https://jour- nals.plos.org/plosone/article?id=10.1371/journal.pone.0222907
3. Nitish Srivastava nitish@cs.toronto.edu Geoffrey Hinton hinton@cs.toronto.edu Alex Krizhevsky kriz@cs.toronto.edu Ilya Sutskever ilya@cs.toronto.edu Ruslan Salakhutdinov rsalakhu@cs.toronto.edu ," Dropout: A Simple Way to Prevent Neural Networks from Overfitting"- Journal of Machine Learning Research 15 (2014).
4. Yu Wang, Student Member, IEEE, Miao Liu, Member, IEEE, Jie Yang, Member, IEEE, and Guan Gui, Senior Member, IEEE- " Data-Driven Deep Learning for Automatic Modulation Recognition in Cognitive Radios".
5. SergeyIoffe SIOFFE@GOOGLE.COM ,ChristianSzegedy SZEGEDY@GOOGLE.COM Google, 1600 Amphitheatre Pkwy, Mountain View, CA 94043- " BatchNormalization: Ac- celeratingDeepNetworkTrainingbyReducing InternalCovariateShift".
6. Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning.Battista Big- gioa,b,∗, Fabio Rolia,b aDepartment of Electrical and Electronic Engineering, University of Cagliari, Italy bPluribus One, Cagliari, Italy.
7. Hadeer Ahmed1(&), Issa Traore1, and Sherif Saad21 ECE Department, University of Vic- toria, Victoria, BC, Canada meresger.hs@gmail.com, itraore@ece.uvic.ca 2 School of Com- puter Science, University of Windsor, Windsor, ON, Canada Sherif.SaadAhmed@uwind- sor.ca
8. Gama, J., ˇZliobait˙e, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. 2013. "A Survey on Concept Drift Adaptation". ACM Comput. Surv. 1, 1, Article 1 (January 2013), 35 pages. DOI = 10.1145/0000000.0000000 ,http://doi.acm.org/10.1145/0000000.0000000
9. Edureka - https://www.youtube.com/watch?v=WSbgixdC9g8&t=69s
10. CHUANLONG YIN , YUEFEI ZHU, JINLONG FEI, AND XINZHENG HE State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China Corresponding author: Chuanlong Yin (dragonyincl@163.com) - "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks".
11. J.Schmidhuber,''Deeplearninginneuralnetworks:Anoverview,''Neural Netw., vol. 61, pp. 85–117, Jan. 2015.
12. L. Liu, L. Shao, X. Li, and K. Lu, ''Learning spatio-temporal representations for action recognition: A genetic programming approach,'' IEEE Trans. Cybern., vol. 46, no. 1, pp. 158–170, Jan. 2016.
13. J. Wu, Y. Zhang, and W. Lin, ''Good practices for learning to recognize actions using FV and VLAD,'' IEEE Trans. Cybern., vol. 46, no. 12, pp. 2978–2990, Dec. 2016.
14. "Deep learning in agriculture": A survey Andreas Kamilaris∗, Francesc X. Prenafeta-Boldú Institute for Food and Agricultural Research and Technology (IRTA), Spain-
15. Kamran Kowsari∗, Donald E. Brown§‡, Mojtaba Heidarysafa§, Kiana Jafari Meimandi§, Matthew S. Gerber§‡, and Laura E. Barnes§‡-" Hierarchical Deep Learning for Text Classification" - ∗Department of Computer Science, University of Virginia, Charlottesville, VA, USA § Department of System and Information Engineering, University of Virginia, Charlottesville, VA, USA ‡ Data Science Institute, University of Virginia, Charlottesville, VA, USA {kk7nc, deb, mh4pk, kj6vd, msg8u, lbarnes}@virginia.edu
16. Scikit-learn: Machine learning in python. J. Mach. Learn. Res., 12:2825–2830, November 2011.
17. Franc ̧ois Chollet et al. Keras. https://keras.io, 2015, Date last ac-cessed 23-March-2017.
18. Mohit Sewak, Sanjay K. Sahay∗and Hemant RathoreDepartment of Computer Science and Information System, Birla Institute of Technology and Science, K. K. Birla Goa Campus, NH-17B, By Pass Road, Zuarinagar-403726, Goa, India - "Comparison of Deep Learning and the Classical Machine Learning Algori1thm for the Malware Detection "
19. D. M. W. Powers. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. Journal of Machine Learning Technologies, 2(1):37–63, 2011.
20. R. Tian, L. M. Batten, and S. C. Versteeg -"Function length as a tool for malware classification". In 2008 3rd International Conference on Malicious and Unwanted Software (MALWARE), pages 69–76, October 2008.
21. Haibo He, Yang Bai, E. A. Garcia, and Shutao Li. Adasyn: - "Adaptive synthetic sampling approach for imbalanced learning". In 2008 IEEE InternationalJointConferenceonNeural-Networks, June 2008.
22. Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia Deepika Bansal1, Rita Chhikara2, Kavita Khanna3,Poonam Gupta4 - (https://creativecommons.org/licenses/by-nc-nd/3.0/)
23. www.oasis-brains.org- Dataset
24. 1st ed - Margaret H. Danham,S. Sridhar (2006) - " Data mining, Introductory and Advanced Topics".
25. Patil, T. R., &Sherekar, S. S.(2013)-"Performance analysis of Naive Bayes and J48 classifi- cation algorithm for data classification", InternationalJournal of Computer Science and Ap- plications.
26. "comparison of machine learning algorithms for clinical event prediction" - (riskofcoro- naryheartdisease) Juan-JoseBeunzaa,b,?,EnriquePuertasa,c,EsterGarcía-Ovejeroa,d,Gema- Villalbaa,e, EmiliaCondesa,GerganaKolevaa,CristianHurtadoa,f,ManuelF.Landecho

Mining Association Rules Based on Incremental Computing Of Big Data

Dr .Anil Vasoya, Akash Khandelia, Nishtha Mishra , Sonali Kamat
Department of Information Technology in Thakur College of Engineering and Technology, Mumbai, India

Abstract— **The discovery of association rule has been known to be useful in different sectors such as marketing, decision analysis, and business management. A vital application area of mining association rules is that the market basket analysis, which studies the buying behavior of shoppers by looking for sets of items that are frequently purchased together. With the growing use of record-based databases whose data is continually being added, the need for incremental mining has been called for in recent important applications. New transaction is appended in complex transaction database, and redundant transaction is discarded as time progresses. Extensive data collection is now a hallmark of information technology these days due to rapid increase of data within organizations. The difficult undertaking is the mining of Association administers in enormous databases. An Apriori calculation is usually used to discover from database the successive component sets. Be that as it may, on account of huge database it will be wasteful in light of the fact that it will require more I/O load. Later drawback of the Apriori algorithm is overcome by many algorithms / parallel algorithms (model). The paper analyzes the performance of the algorithm over different datasets and over n processors on a commodity cluster of machines. In the Apriori Algorithm all procedures need to synchronize after each pass. In the event that any procedure is doled out more burden than different procedures in the framework, the slowest procedure will direct the speed of the program. It is therefore important to ensure that load is equally balanced among all processes. Our algorithm decides the no. of running procedures and partitions, the load similarly in order to amplify the framework execution and its proficiency. Utilizing gradual figuring here suppliers a bit of leeway that the prepared information isn't handled again and just the new information entered is prepared each time which is less tedious. In this research paper we are going to use Distributed and Parallel association mining on big data while using Incremental mining with an aim to process a big chunk of data simultaneously in a short span of time.**

***Index Terms — Mining Association Rules, Incremental Computing, Big data, Discovering Large Itemset, Apriori Algorithm, Distributed/Parallel Algorithm, Multithreading*.**

I. INTRODUCTION

To remove valuable and noteworthy data from huge information, the data innovation (IT) world is confronting enormous issues today. At present the information existing is in Tera or peta bytes and should increment to peta or zetta bytes in close by future in different organizations. Examples of such applications incorporate Web log records, financial exchange information; basic food item deals information, exchanges in electronic business, and every day climate/traffic records, to give some examples. Such information can be as organized, unstructured and semi-organized datasets.

Such information must be mined to get learned example from it and accordingly information mining methods and calculations are utilized.

Data Mining is a lot of technique that applies to huge and complex databases. This is to take out the haphazardness and find the concealed example. As this information mining techniques are quite often computationally escalated. We use information mining devices, procedures, and speculations for uncovering designs in information. There are excesses of main impetuses present. Also, this is the motivation behind why Data mining has become such a significant zone of study.

An Association rule assumes a significant job in late information mining procedures. The acquiring of one item alongside another related item speaks to an affiliation rule. Association rules are utilized to show the connections between information things. Association rules are as often as possible utilized for various purposes like showcasing, promoting and stock bazaar. Association precludes discover normal use of things.

This issue is propelled by applications known as market container investigation to discover connections between things obtained by clients. Databases may accumulate an unlimited quantity of knowledge to be mined. Mining association rules into such databases might involve significant processing power.

II. BACKGROUND

Association Rule Learning may be a general technique accustomed discovers associations amongst numerous variables. It's often utilized by grocery stores, retailers, and anyone with a bulky transactional database. Association rules are if/then statements that protection to get associations between apparently unrelated data in an exceedingly electronic information service or other information storehouse. An example of an association rule would be "If a customer buys a dozen breads, he is 80% likely to also purchase butter/jam." Association rules are shaped by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important associations. Support is an indication of how frequently the items emerge in the database.

In data processing, association rules are helpful for analyzing and predicting customer nature. They play a significant role in shopping basket data analysis, item clustering, catalog design. Programmers use association rules to construct programs of machine learning. Machine learning could be a form of computing that seeks to assemble programs with the potential to turn into more competent without being explicitly programmed.

III. RELATED WORK

In this section, we review some related work including frequent itemsets, incremental mining of frequent itemsets.

3.1 FREQUENT ITEM SETS

Let I = {i1, i2… in} be the set of all items. An itemset X is a nonempty subset of I. For brevity, an itemset can be denoted as X = {ij1, ij2…ijn} where 1<= j1, <j2,…<=n An itemset with k items is called a k-itemset. The duple (tid, X) is known as an exchange if tid is an exchange identifier and X is an item set. An exchange database TDB is a lot of exchanges, and |TDB| is the quantity of exchanges. An item set X is contained in exchange (tid, Y). Given an exchange database TDB, the recurrence of an item set X, indicated recurrence (X), is the quantity of exchanges in TDB that contain X. The help of an item set X, signified support (X), is given by

Support (X) = Frequency (X) / |TDB|

There are two kinds of item sets. The principal comprises of item sets that express just a solitary Boolean connection, either true or false, for example, {(butter, true), (chesses, true)}. For comfort, we can speak to this as {butter, cheese}, implying that butter and chesses were obtained together. Conversely, if an item set has a false connection {(butter, false), (chesses, false)}, it shows that these items were not acquired. In the subsequent kind, everything is a property estimation pair. Each characteristic may have different qualities, for example, {(age, youthful), (sex, male)}, and {(age, old), (sex, female)}. Here age and sexual orientation are two traits, (youthful, old) and (male, female) are their qualities.

3.2 INCREMENTAL MINING OF FREQUENT ITEM SET

The undertaking of steadily mining and refreshing successive item sets has come to speak to a significant field of research. After a database is refreshed, some current examples will probably never again be significant, and some new examples might be presented. In 1996, Cheung et al proposed the FUP calculation for productive upkeep of found affiliation rules when new information is added to an exchange database. FUP depends on Apriori and is utilized to find the new regular itemsets in a powerful database. The idea is to store the number of occurrences of the frequent itemsets found in a previous mining operation. By using these stored counts and scanning the newly added transactions, the overall count of the new candidate itemsets is eventually obtained. Leave dB alone a lot of new exchanges and TDB be the refreshed database (counting all exchanges from TDB and dB).An Itemset X is either visit or inconsistent in TDB or db. In this manner, there are four cases, as appeared in Table 1.In its first cycle, FUP filters db to acquire the quantity of events of every 1-itemset. Since the event includes of all successive itemsets in TDB are known ahead of time, the absolute number of events for a subjective X is effortlessly determined if X and db fall into case 2. In the terrible circumstance of case 3, TDB must be rescanned. Also, the following pass checks db to include the up-and-comer 2-itemsets in db. In the event that fundamental, TDB will be again rescanned on the off chance that 3. The procedure is rehashed until all continuous itemsets have been found. In the most pessimistic scenario, FUP doesn't lessen the quantity of outputs of the first database.

XinTDB/Xindb	Frequent Itemset	Infrequent itemset
Frequent Itemset	Case 1: frequent	Case 2: be easily calculated
Infrequent Itemset	Case 3 rescan the TBD	Case 4: infrequent

Table 1 Possible case for updating with FUP

In 1997, Cheung et al depicted the FUP2 calculation, which is a more broad gradual procedure than FUP. FUP2 is effective not just on the assignment of building up a database yet additionally on cutting information. Tomas et al proposed another strategy to quicken incremental mining, by keeping up a negative border. In 1999, Ayan et al proposed the Overhaul with Early Pruning (UWEP) strategy, which utilizes an energetic lookahead methodology to up grade existing visit itemsets by recognizing and expelling item sets that are occasional within the upgraded database. In 2001, Lee et al proposed the Sliding Window Sifting (SWF) approach. SWF allotments databases and applies a sifting edge to each segment to produce candidate item sets. In 2002, Veloso et al depicted the Crisscross calculation, which employments a list of tid's and computes maximally visit item sets within the upgraded database to maintain a strategic distance from producing various pointless candidates.

To overcome the destitute execution that results from rescanning the initial database, Chang et al proposed the NFUP calculation (''New FUP'') for effective incremental mining of association rules from huge exchange databases. NFUP could be an in

reverse strategy that as it were requires scanning the incremental database.

Rather than rescanning the initial database for a few recently created visit item sets within the incremental database,it checks the events of recently produced visit it emsets and essentially erases clearly occasional ones.

IV. QUICKLY UPDATE METHOD

The method of incremental mining is outlined in Fig. 1. Incremental mining is executed by utilizing the initial database, Incremental database and Unique Patterns/Rules as information source. The first database is continuously gigantic; in this manner, incremental mining conveys more dependable execution by extricating rules from the incremental database and combining them with the initial rules without filtering the initial database. In any case, a few calculations proposed within the past as it were decrease the number of times the database is checked instead of mining without checking the database.

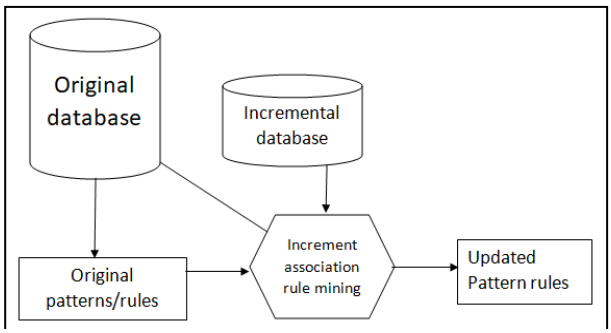


Fig 1: Process of incremental mining

Previously proposed calculations continuously have a back number, which may be an edge to choose visit itemsets. Be that as it may, agreeing to the long tail hypothesis total things that as it were show up rarely may create more benefit than all visit things together. The utilize of a bolster number set progresses productivity, but too creates issues, such as the issue incremental mining. With the coming of enormous information;the sum of modern exchange inf ormation has been expanding relentlessly, and rare things would be killed by to begin with developing a tree. Indeed in the event that these things were too ended up visit by the expansion of ensuing information, they would not be considered. In this way, the execution of incremental mining is required each time; i.e., all the information needs to be re-scanned, which lowers the performance.

This paper proposes the Rapidly Upgrade Strategy (QUM), which eradicates the conventional least bolster number to extend the adaptability of incremental affiliation run the show mining.

V. DATA SETS

The experimentation is carried out with the help of synthetic datasets that are generated through the use of a dataset generator that is publicly available. A data set is a gathering of data, frequently presented in tabular shape. Each column represents a particular variable. Each row corresponds to a specified associate of the data set in question. It gives values to every variable, such as transaction id and transaction of an object. Each value is known as a element. The data set might consist of data for one or more members, equivalent to the number of rows.

The Dataset is anonymzed and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users distributed across 6 csv files. Now let's jump straight into the data and do some exploratory analysis. Let's first have a look at these files:

Orders: This file gives a list of all orders we have in the dataset. One row per order. The orders.csv doesn't tell us about which products were ordered. This is contained in the order_products.csv

order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
2539329	1	prior	1	2	8	NA
2398795	1	prior	2	3	7	15
473747	1	prior	3	3	12	21
2254736	1	prior	4	4	7	29

Order_product_train: This file gives us information about which products (product_id) were ordered. It also contains information of the order (add_to_cart_order) in which the products were put into the cart and information of whether this product is a re-order (1) or not (0).

Order_id	Product_id	add_to_cart_order	reordered
1	49302	1	1
1	11109	2	1
1	10246	3	0
1	49683	4	0

Products: This file contains the names of the products with their corresponding product_id. Furthermore the aisle and department are included.

Product_id	Product_name	Aisle_id	Department_id
1	Chocolate sandwich cookies	61	19
2	All Seasons Salt	104	13
3	Robust Golden Unsweetened Oolong Tea	94	17
4	Smart Ones Classic Favorites Mini Rigatoni With Vodka cream Sauce	38	1

Order_products_prior: This file is structurally the same as the other_products_train.csv.

Order_id	Product_id	add_to_cart_order	reordered
2	33120	1	1
2	28985	2	1
2	9327	3	0
2	45918	4	1

Aisles: This file contains the different aisles.

Aisle_id	Aisle
1	prepared soup salad
2	Speciality cheese
3	energy granio bars
4	instant foods

Departments:

department_id	department
1	frozen
2	other
3	bakery
4	produce

VI. PROPOSED WORK

6.1 ASSOCIATION RULE MINING ALGORITHM

A rule of association implies definite interaction between a set of objects within a database. An association rule is an expression of the form A, B, where the elements are A and B. The observant logic of such a concept is that database transactions containing A are likely to contain B. Association rule is one of the data starting databases which can be used by a company decision maker to build and contribute to its overall earning. Agrawal et al. first established the rules on the mining of associations in databases in 1993. They considered the following example; 60 percent of the translations buying bread are also transactions buying milk.

The formal statement is as follows. Let I = {i1, i2… im} represent the set of literals, called items. The symbol T represents an arbitrary transaction, which is a set of items (Item set) such that T ⊆ I. Each transaction has a unique identifier, TID. Let D be a database of transactions. Assume X is an Item set; a transaction T contains X if and only if X ⊆T. An association rule applies in the form X→ Y, where X ⊆I, Y ⊆I and X ∩Y = (For example, (I = {ABCDE}, X = {AC}, Y = {BE}). An association rule X →Y, has two properties, support and confidence. When s% of transactions in D contains XU Y, the support of the rule X→ Y is s%. If some of the transactions in D contain X and, c% also contain Y, then the confidence in the rule X→ Y is c%. In general, the confidence is expressed in the form confidence(X → Y) = support (XUY)/support(X). Given the user-assigned minimum support (minSup) and minimum confidence (minConf) thresholds for the transaction database D, here the rules on the mining of associations are to consider all rules whose support and confidence surpass the two thresholds respectively.

6.2 INCREMENTAL ASSOCIATION RULE MINING

The mining of transactional database association rules is typically an offline process, as finding the association rules in large database is costly. New transactions are created with normal market-basket applications and old transactions may become obsolete as time progresses. As a result, incremental updating techniques should be developed to preserve the discovered association rules so that mining on the entire modified database is not redone. A database may allow frequent or occasional updates, and such updates may not only invalidate existing rules on associations but also activate new rules. Thus it is nontrivial to maintain such discovered rules in large databases. Considering an original database and newly inserted transactions, the following four cases may arise:

Case 1: An itemset is large in the original database and in the newly inserted transactions.

Case 2: An itemset is large in the original database, but is not large in the newly inserted transactions.

Case 3: An itemset is not large in the original database, but is large in the newly inserted transactions.

Case 4: An itemset is not large in the original database and in the newly inserted transactions.

Since itemsets in Case 1 are large in both the original database and the new transactions, they will still be large after the weighted average of the counts. Similarly, itemsets in Case 4 will still be small after the new transactions are inserted. Thus Cases 1 and 4 will not affect the final association rules. It could be a possibility that Case 2 may remove existing association rules, and Case 3 may add new association rules. A good rule-maintenance algorithm should thus accomplish the following:

1. Asses large set of objects in the original database and decide if are still significant in the revised database.
2. In the modified database, find out if any tiny itemsets in the original database can get huge.
3. Search for itemsets that only appear in newly inserted transactions, & determine if they are big in updated database.

6.3 DISCOVERING LARGE ITEMSETS

Algorithms for discovering large itemsets make multiple passes over the data. In the first pass, we count the support of individual items and determine which of them are large, i.e. have minimum support. In each subsequent pass, we start with a seed set of itemsets found to be large in the previous pass. We use this seed set for generating new potentially large itemsets, called candidate itemsets, and count the actual support for these candidate itemsets during the pass over the data. At the end of the pass, we determine which of the candidate itemsets are actually large, and they become the seed for the next pass. This process continues until no new large itemsets are found.

6.3.1 APRIORI ALGORITHM

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceed by identifying the recurring individual items in the database as well as extending them to bigger item sets as long as those item sets come out adequately often in the database. The frequent item sets examined by Apriori and can be used to create a conclusion association rules which depict interest to general trends in the database.

Figure 1 gives an overview of the Apriori Algorithm,

using the notation in Table 1. The first pass of the algorithm simply counts item occurrences to determine the large 1-item-sets. A subsequent pass, say pass k, consists of two phases. First, the large item-sets L_{k-1} found in the $(k-1)^{th}$ pass are used to generate the candidate item-sets C_k. Next, the dataset is scanned and the support of candidates in C_k is counted. For fast counting, we need to efficiently determine the candidates in C_k that are contained in a given transaction t.

K itemset	An item-set having k items
L_k	Set of frequent k itemsets (with minimum support)
C_k	Set of candidate k itemsets (potentially frequent)
P^i	Processor with id i
D^i	The database local to the processor P^i
C^i_k	Candidate set maintained with the Processor P^i

Table 2 Notations

```
1)  L1 = { Large 1-item-sets};
2)  k =2; // k represents the pass number
3)  while ( L_{k-1} <> 0 ) do
4)  begin
5)      C_k = new candidates of size k generated from L_{k-1};
6)      forall transactions t ϵ D do
7)          Increment the count of all candidates in C_k that a
    contained in t;
8)      L_k = All candidates in C_k with minimum support;
9)      k = k +1;
10) end
11) Answer = U_k L_k;
```

Fig 2: Apriori Algorithm

6.4 DISRIBUTED/ PARALLEL COMPUTING

Database will generate huge amount of data that need to be processed. Rules relating to mining association in such databases may involve significant processing power. A distributed network may be a potential solution to that problem. Moreover, lots of databases are distributed in nature which may assemble it more possible to use distributed algorithms. Moreover, lots of databases are distributed in nature which may assemble it more possible to use distributed algorithms. Calculation of the collection of large item sets in the database is the key structure of mining association rules. Distributed computation of large sets of objects faces many new problems. Obviously, one can measure large article sets locally, but a collection of locally large items may not be globally large. Since it is very costly to transmit the whole data set to other sites, one alternative is to put on air all the counts of all the item sets. Nonetheless, a database may take hold of very large combinations of item sets, and a huge amount of communication will be engaged for the moment.

VII. RESULT AND DISCUSSION

7.1 HORIZONTAL ALGORITHM

Let D be a transaction database. We view D as binary matrix of N rows and L columns, where each row is a transaction over some set of items, A = {a1... aL}, and each column represents one of the items in A.

The database D is partitioned horizontally between M players, denoted P1,...,PM. Player Pm holds the partial database Dm that contains Nm = |Dm| of the transactions in D, $1 \le m \le M$. The unified database is D = D1 ∪⋯∪ DM, and it includes N: = $\sum_{m=1}^{M} N_m$ transactions. An itemset X is a subset of A. Its global support, supp(X), is the number of transactions in D that contain it. Its local Support, $supp_m(X)$, is the number of transactions in Dm that Contained. Clearly, $supp(X) = \sum_{m=1}^{M} supp_m(X)$. Let s be a Real number between 0 and 1 that stands for a required support threshold. An itemset X is called s-frequent if $supp(X) \ge sN$. It is called locally s-frequent at Dm if $suppm(X) \ge sNm$.

For each $1 \le k \le L$, let F^k_s denote the set of all k-itemsets (namely, itemsets of size k) that are s-frequent, and $F^{k,ms}$ be the set of all k-itemsets that are locally s-frequent at Dm, $1 \le m \le M$. Our main computational goal is to find, for a given threshold support $0 < s \le 1$, the set of all s-frequent itemsets, Fs: = $U^L_{k=1} F^k s$. We may then continue to find all (s, c) association rules, i.e., all association rules of support at least sN and confidence at least c. (Recall that if X and Y are two disjoint subsets of A, the support of the corresponding association rule $X \Rightarrow Y$ is supp(X ∪ Y) and its confidence is supp(X ∪ Y)/supp(X).)

The Algorithm proceeds as follows:

(1) **Initialization:** It is assumed that the players have already jointly calculated $F^{k-1}{}_s$. The goal is to proceed and calculate $F^k{}_s$.

(2) **Candidate Sets Generation:** Each player Pm computes the set of all $(k-1)$-itemsets that are locally frequent in his site and also globally frequent; namely, Pm computes the set $F^{k-1,m}{}_s \cap F^{k-1}{}_s$. He then applies on that set the Apriori algorithm in order to generate the set $B^{k,m}{}_s$ of candidate k-itemsets.

(3) **Local Pruning:** For each $X \in B^{k,m}{}_s$, Pm computes suppm(X). He then retains only those itemsets that are locally s-frequent. We denote this collection of itemsets by $C^{k,m}{}_s$.

(4) **Unifying the candidate itemsets**: Each player broadcasts his $C^{k,m}{}_s$ and then all players compute $C^k{}_s$: = $U^M{}_{m=1} * C^{k,m}{}_s$.

(5) **Computing local supports**: All players compute the local supports of all itemsets in $C^k{}_s$.

(6) **Broadcast Mining Results**: Each player broadcasts the local supports that he computed. From that, everyone can compute the global support of every

itemset in C^k_s. Finally, is the subset of C^k_s that consists of all globally s-frequent k-itemsets.

In the first iteration, when k = 1, the set $C^{k,m}_s$ that the mth

Player computes (Steps 2-3) is just F^{k-1}_s, namely, the set of single items that are s-frequent in Dm. The complete FDM algorithm starts by finding all single items that are globally s-frequent. It then proceeds to find all 2-itemsets that are globally s-frequent, and so forth, until it finds the longest globally s-frequent itemsets. If the length of such itemsets is K, then in the (K +1)th iteration of the FDM it will find no (K + 1)-itemsets that are globally s-frequent, in which case it terminates.

A **running example**:

Let D be a database of N = 18 itemsets over a set of L = 5 items, A = {1, 2, 3, 4, 5}. It is partitioned between M = 3 players and the corresponding partial databases are:

D1 = {12, 12345, 124, 1245, 14, 145, 235, 24, 24}

D2 = {1234, 134, 23, 234, 2345}

D3 = {1234, 124, 134, 23}

For example, D1 includes N1 = 9 transactions, the third of which (in lexicographic order) consists of 3 items - 1, 2, and 4. Setting s = 1/3, an itemset is s-frequent in D if it is supported by at least 6 = sN of its transactions. In this case, F^1_s = {1, 2, 3, 4; F^2_s = {12, 14, 23, 24, 34; F^3_s = {124}; $F^4_s = F^5_s = \varnothing$,

And $Fs = F^1_s \cup F^2_s \cup F^3_s$. For example, the itemset 34 is indeed globally s-frequent since it is contained in 7 transactions of D. However, it is locally s-frequent only in D2 and D3.

In the first round of the FDM algorithm, the three players compute the sets $C^{k,m}_s$ of all 1-itemsets that are locally frequent at their partial databases:

$C^{1,1}_s$ = {1, 2, 4, 5}; $C^{1,2}_s$ = {1, 2, 3, 4}; $C^{1,3}_s$ = {1, 2, 3, 4}.

(Note that 15, 25, 45 are locally s-frequent at D1 but they are not included in $C^{2,1}_s$ since 5 was already found to be globally infrequent.) Hence, C^2_s = {12, 13, 14, 23, 24, 34}.

Then, after verifying global frequency, we are left with F^2_s = {12, 14, 23, 24, 34}

In the third round, the candidate itemsets are:

$C^{3,1}_s$ = {124}; $C^{3,2}_s$ = {234}; $C^{3,3}_s$ = {124}.

So, C^3_s = {124, 234} and, then, F^3_s = {124}. There are no more frequent itemsets.

7.2 VERTICAL ALGORITHM

First we use Partition algorithm(PAFI) to find clusters, th en we use Enhanced Apriori algorithm to find the commo nly used itemsets. Steps as followed:

(1) For a given set of transactions in the database D, it applies partition algorithm in order to find clusters based on the number of transactions. Here we are getting 2 clusters CL1 and CL2.

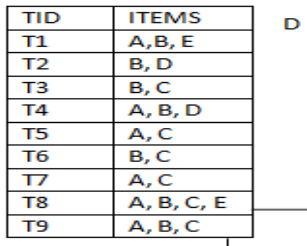
(2) Here CL2 has less number of transactions that is less than the threshold value so we are deleting the transactions in CL2 and we are concentrating on the transactions in CL1.

(3) Now apply the improved Apriori algorithm, by finding the count of an each item from D1.Since scanning is not done using the entire database transactions D it improves the efficiency. These set of items will be considered as candidates C1.These transactions will be considered as D1.

(4) D1 is scanned to generate C2 candidates and find the count of each candidate.

(5) Compare candidate support with minimum support. The candidates having less count than the minimum

(6) Support will be deleted. The above process is repeated for C3. This will be generated for Ck until Ck+1 becomes empty



D

TID	ITEMS
T1	A,B, E
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

CL1

TID	Items
T8	A,B,C,E
T1	A,B,E
T3	B,C
T5	A,C
T6	B,C
T7	A,C
T9	A,B,C

CL2

TID	Items
T4	A,B,D
T2	B,D

D1

Items	count
A	5
B	5
C	6
E	2

C1, l1

TID	Items
T8	A,B,C,E
T1	A,B,E
T3	B,C
T5	A,C
T6	B,C
T7	A,C
T9	A,B,C

Itemset	count
{A,B}	3
{A,C}	4
{A,E}	2
{B,C}	4
{B,E}	2
{C,E}	1

C2

L2

Itemset	count
{A,B}	3
{A,C}	4
{A,E}	2
{B,C}	4
{B,E}	2
{C,E}	1

L3

L4

Itemset	count
{A,B,C,E}	1

VIII. CONCLUSION

In this paper, Association mining is proposed along with the improvements needed in the algorithm to make in efficient and better when it comes to large data sets. Here we see the usage of incremental mining by using Horizontal algorithm and Vertical algorithm on structured data. Where incremental mining helps us to avoid processing the already processed data again and saves the processed data at a location which is the used in future to find FIS while the new data is added, this gives us an advantage over other algorithms where the whole data is processed again example in Hadoop where MapReduce is used for this. This method is useful for small or medium datasets but not for big data analysis as it becomes time consuming. Here this research paper gives a brief idea of the results that can be achieved by using Partitioning and Distributed computing in Incremental mining using big chunks of data. The results and discussion shows us the steps carried out in both horizontal and vertical algorithms respectively.

References

[1] Anil Vasoya, Dr. Nitin Koli, "*Mining of association rules on large database using distributed and parallel computing*", 7th International Conference on Communication, Computing and Virtualization, 2016

[2] Neha Saxena, Rakhi Arora, Ranjana Sikarwar, Ashika Gupta, "*An Efficient Approach of Association Rule Mining on Distributed Database Algorithm"*, International Journal of Information and Computation Technology, 2013

[3] Yao Li, Zhi-Heng Zhang, Wen-Bin Chen, Fan Min, "*An approach to incremental mining of frequent itemsets with three-way-decision pattern updating"*, Springer-Verlag Berlin Heidelberg, 2015

[4] Hong-Yi Chang, Jia-Chi Lin, Mei-Li Cheng, Shih-Chang Huang, "*A Novel Incremental Data Mining Algorithm based on FP-Growth for Big Data*", International Conference on Networking and Network Applications, 2016

[5] Toshi Chandraker, Neelabh Sao, "Incremental Mining on Association Rules", International Journal of Engineering and Science, 2013

[6] Rakesh Agrawal, Ramakrishnan, "*Fast Algorithms for Mining Association Rules*", Proceedings of the 20th VLDB Conference, 1994

[7] Ketan D. Shah, Dr. (Mrs.) Sunita Mahajan, "*Maximizing the Efficiency of Parallel Apriori Algorithm",* International Conference on Advances in Recent Technologies in Communication and Computing, 2019

[8] Tamir Tassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", IEEE Transactions on Knowledge and Data Engineering, 2013

[9] D. Keranahanirex, Dr M.A. DoraiRangaswamy, "*Efficient Algorithm for Mining Frequent Itemsets Using Clustering* Techniques", International Journal on Computer Science and Engineering, 2011

[10] Sussan Einakian, M. Ghanbari, "*Parallel Implementation of Association Rule in Data Mining",* Proceedings of the 38th Southeastern Symposium on System Theory Tennessee Technological University Cookeville, 2016

[11] Subbulakshmi. B, Monisha. M, "*Incremental Constraint Class Association Rule Mining of Student Performance Dataset",* Fifth International Conference on recent trends in Information Technology, 2016

[12] Sagar Bhise, Prof. Sweta Kale, "*Efficient Algorithms to find Frequent Itemset Using Data Mining*", International Research Journal of Engineering and Technology (IRJET), 2017

[13] Diana-Lucia Miholcaa, Gabriela Czibula, Liana Maria Criveia, "*A new incremental relational association rules mining approach",* International Conference on Knowledge Based and Intelligent Information and Engineering, 2018

[14] Wenliang Cao, Xuanzi Hu, Chongjie Dong, "*Research of the Mining Algorithm Based on Distributed Database*", Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2011

Food Ordering with AR

Nikita Badod, Yojana Bavkar, Darpan Jain, Neha Kapadia
Thakur College of Engineering and Technology, Mumbai, India
nbadod23@gmail.com, yojanabavkar@gmail.com, jaindarpan11@gmail.com, neha.kapadia@thakureducation.org

Abstract—***Restaurants, nowadays, have become a prime business sector. In the last decade there has been a 21% growth in the new restaurants, setup around India itself. With the rising competition in market every food outlet craves a need to stand unique to attract customers. One of the key problems is the waiting system in restaurants during holiday seasons. The procedure can be advanced by the use of digital menus. The track records of the customers and their bills is also manually handled. This gives exposure to errors. This can be easily tackled by keeping a simple database. The strategy is to assign a tablet to each table with restaurant's application installed in it. The app not only enables the software to serve multiple needs of the customers, but also save their time.***

Keywords—order food, tablet, application, augmented reality

I. INTRODUCTION

A. Food is a very basic need of every animal on this earth and humans are no exception to it. Along with food, the magic of taste has been developed among humans, leading to the overwhelming business of restaurants across the world. Restaurants are a prime option taken into consideration while planning an outing. Hence, there is a lot of chaos and waiting during holiday season.

B. In the age of Modernization, technological advances have made lives faster, easier and better. Food ordering using AR is a similar attempt at enhancing the customer experience as well as digitalizing the food ordering protocol. The food ordering system is an internet based application utilized in mobile for table reservation in addition to food ordering. Every table is equipped with a tablet. Customers can either login or skip the login. They can order and check bill through the application. The restaurant staff (along with the receptionist and manager), will be connected on app. They will receive order and can accordingly prepare food, deliver it and accept payment. In the subsequent upgrades of the application, user experience will be enhanced.

II. BACKGROUND

When a customer arrives into the restaurant, a table is allotted. This table already has a tablet on it with an application running. The customer is provided with two choices, either to register or skip the registration, After the desired option is selected, then accordingly either a known session is created or a random session is created. After this a menu is appeared on the screen. The customer can browse through the menu and order accordingly. After the order is placed the current order statuses are displayed. Also, the data is being stored in the database on the backend simultaneously. According to the order, a bill is also generated. The application further asks whether the payment be done through cash or card. The app is based on the very idea of development of the business. The increase in the quality of a user experience through the application should be remarkable.

III. RELATED WORK

Kunal Raut, Priyanka Khare, Aishwarya Kamble, Sachin Deshpande's Augmented Reality in Restaurant helps us to understand different techniques to use AR in a restaurant. This method can fancy a restaurant and provide a real-time experience. Their research helps us know about the ongoing or finished projects into the similar stream. The paper suggested various features including "I am feeling lucky" which is based on environmental factors. Paper is mainly focused on to the AR part and it hence doesn't provide a customized app for a particular restaurant. It also doesn't include a system of proper food ordering and paying bills.[1]

Ashwini J, Ayisha Shetty, Chaithra, Lasya K Rao, Prof. Ramya D Shetty's Automated Food Ordering System identified the problems with the traditional food ordering method. They have highlighted that the most common blunder is waiters making mistakes with customer's orders. At times, a waiter can forget to add a specific item, make a change because a customer is allergic to certain substance, or forget to give the order to the kitchen. Impatient customers also call over the waiter/waitress frequently to find out the status of their order several times during their visit, wasting the waiter's service time. Customers have to wait for a waiter to take their order. They must rely on the waiter to remember their order and specific details. Their food may take longer to be prepared and served if the waiter has multiple tables. They may also get wrong bills since they cannot see their bill amount until their meal is complete. Waiters need to constantly check with the chefs to determine when food is ready. Conversely, chef needs to make sure waiters know that food is ready. Managers have to analyze hundreds of paper receipts to determine best-selling items, popular hours and customer satisfaction. They also require re-printing of menus when food is not available or a price needs to be changed. This can be costly and time-consuming to a restaurant.[13]

Adithya R, Abhishek Singh, Salma Pathan, Vaishnav Kanade's Online Food Ordering System proposes flexibility to the Customers/Users to order from either Restaurants or Mess. It will also provide Recommendations to the customers from the restaurants/mess owners uploaded on a daily basis. In the proposed system, there will be no limitation on the amount of order the customer wants. Also, same application can be used as a Startup Business for the

developers. It will provide real time customers feedback and ratings along with the comments to the restaurants/mess owner. It gives appropriate feedbacks to users, so if there is any error happened, then there will be a feedback dialog toward users. Online Food Ordering System is based on Internet of Things. It is a wireless food ordering system using android devices. Android devices have gained immense popularity and have revolutionized the use of mobile technology in the automation of routine task in wireless environment. Android is a Linux based operating system for mobile devices such as smartphones and tablets. Their objective is to design a system that is able to accommodate huge amount of orders at a time and automatically compute the bill. They have focused performance and acceptability in terms of security, user-friendliness, accuracy and reliability and to improve the communication between the client and customers is one of the objectives.[14]

IV. PROBLEM DEFINITION

The ambience in the restaurants can be improvised with use of digitalized menus. It also provides efficient working environment with minimal errors. The need for better facilities to the customer is the prime focus of the client. The proposed system, which includes Augmented Reality, is an upgrade, every restaurant deserves.

The purpose of this project is to develop a Food Ordering System which not only orders food but also helps in reserving table. So instead of waiting for the waiter to take orders one can look into the app and order food. Once the customer login it will redirect the customer to the application. The app will provide the Menu of the restaurant along with which the table no will also be displayed. Visualization of menu items using AR and provide the customers free entertainment facilities. It becomes difficult to manage restaurant during peak hours due to unavailability of waiters. Traffic Analysis needs to be done during long waiting and tables start filling simultaneously. Many people ordering at the same time can create some traffic on server. In such cases allotment of order preference needs to be done properly so the customer receive order in time. Along with this, service efficiency and effectiveness in the shortest possible time is expected by clients in the restaurant sector. As the system is online based, provision of internet through Wi-Fi should be done. Orders should be regularly and properly updated in the staff's portal/account. The procedure can be optimized with the help of internet and technology. It will help customers build a new experience of tracking order and provide entertainment while the order gets ready. There should be a way to recognize regular customers. Keeping a database can help in resolving this issue.

The Food ordering system serves many purposes, including reservation of a table of the respective restaurant. The main objective of the application is to optimize the food ordering and waiting system in a restaurant. Restaurants or food outlets at a mall to digitalize the waiting system and reduce labor can use this application. It can reserve a table, record user information, order food, display order status, inform the waiter assigned, entertain, generate bill and various other features. Another objective of this project is to provide better customer experience and make a good impression by a sophisticated system. Customer's personal information will be recorded, which can help the restaurants to analyze their business. There is a scope of lot of upgrades like games to play or books to read can be provided in the application for users until the order arrives.

V. OBJECTIVES

A. The app is based on the very idea of development of the business. The increase in the quality of user experience through the application should be remarkable. The customer has an easy control on what to order which also overcomes the communication barrier errors. The database maintain helps understand the likes and dislikes of the customers. A feedback form in the application can help in understanding their comforts and expectations. The digitization also helps in modernizing the ambience of the restaurant. The contact and emails of user can be used to notify them about the upcoming festive offers.

B. All the changes made in the applications are done for the comfort of customers. The customers play a vital role in this project. When a restaurant agrees to digitalize their menus, they put a heavy trust over its smooth working. The app will be carrying out updates based on customer responses which again favors the customer. This importance given to client will surely make one feel dazzled. While the order is being prepared, some game or recreational will be provided through app. The application will be interactive. There will be personal and better offers for regular customers.

VI. METHODOLOGY

The methodology that we chose for our project is Agile, while waterfall model is a tried and tested framework, it is not suitable for our project, due to the changeable nature and small team size. Agile allows us to be flexible, manage changing requirements, manage the ever-increasing scope as well as get consumer perspective. Agile has focus on customer satisfaction, which is a crucial aspect of any healthcare application. Patients need to be clued in to the process for the indispensable inputs they can provide.

Agile software development is a methodology for software projects under whose framework project requirements and solutions evolve through the collaborative effort of self-organizing and cross-functional teams and their customer and intended users. It advocates for flexible modelling and planning, flexible implementation, rapid delivery of code, and constant improvement, and it advocates speedy and flexible reactions to change.

The 12 agile principles are -

Our highest priority is to satisfy the customer through early and continuous delivery of valuable

software.

1. Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
2. Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
3. Business people and developers must work together daily throughout the project.
4. Build projects around motivated individuals. Give them the environment and support they need and trust them to get the job done.
5. The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
6. Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
7. Continuous attention to technical excellence and good design enhances agility.
8. Simplicity - The art of maximizing the amount of work not done--is essential.
9. The best architectures, requirements, and designs emerge from self-organizing teams.
10. At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly

VII. SYSTEM DESIGN

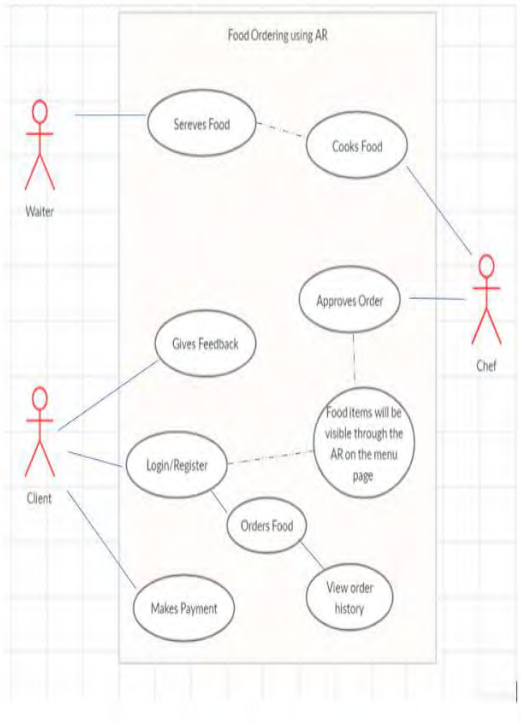


Fig.1 Use case Diagram

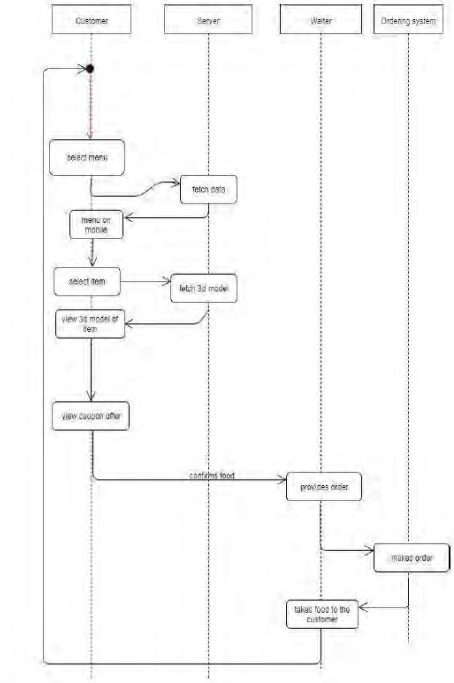


Fig.2 Sequence Diagram

VIII. RESULTS & DISCUSSIONS

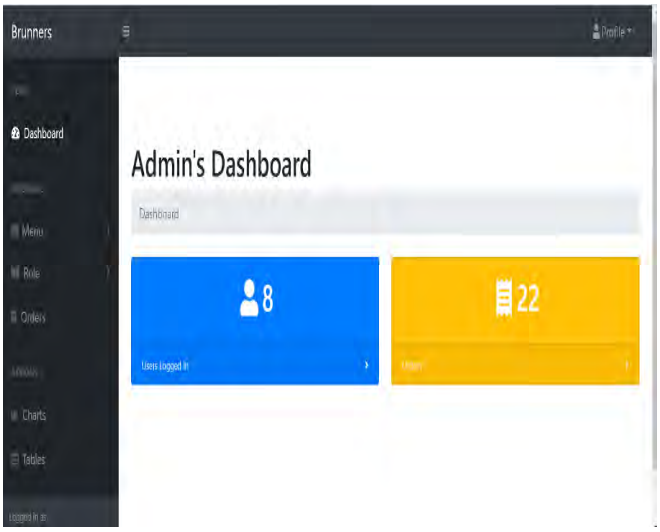


Fig.3 Admin Dashboard

The admin dashboard comprises of no. of customers logged in, total no. of orders received, no. of tables reserved.

It handles the following portal:

- Receptionist
- Customer
- Chef
- Waiter

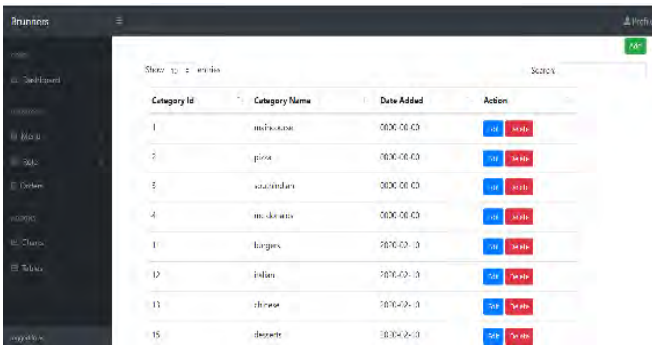


Fig.4 Category Page

The interface Menu comprises of Category and Product Page. In this section, Admin can Add, Update or Delete restaurant categories and products.

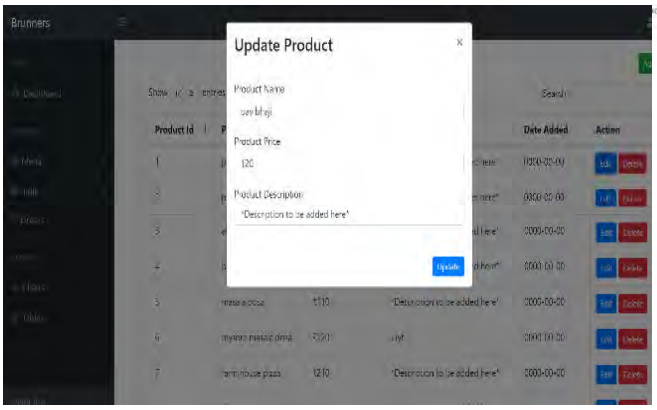


Fig.5 Update Product Page

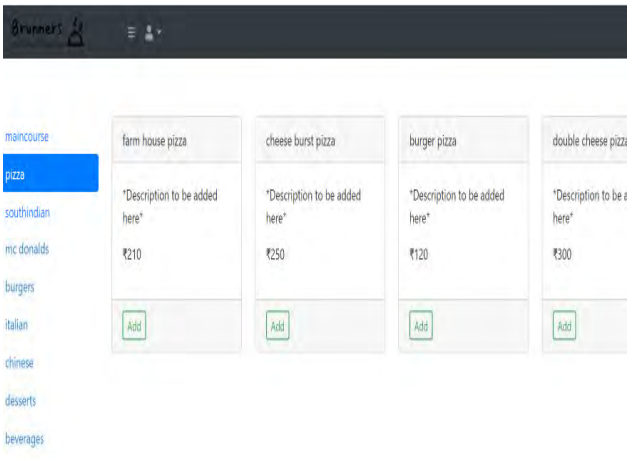


Fig.6 Menu Page

The Menu page gives information about the menu items which will be visible to the customers who have logged in through this website.

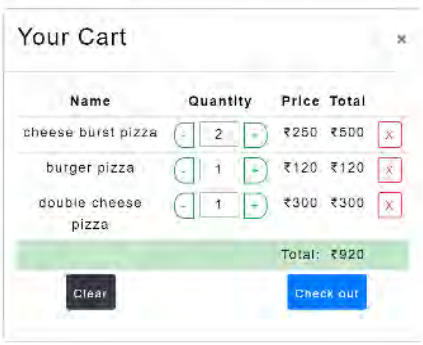


Fig.7 Cart Page

The cart page consists of the items which customers wishes to order. It has been designed with Ajax technology. The customer can view, delete or clear the product items.

Fig.8 Delete Cart Page

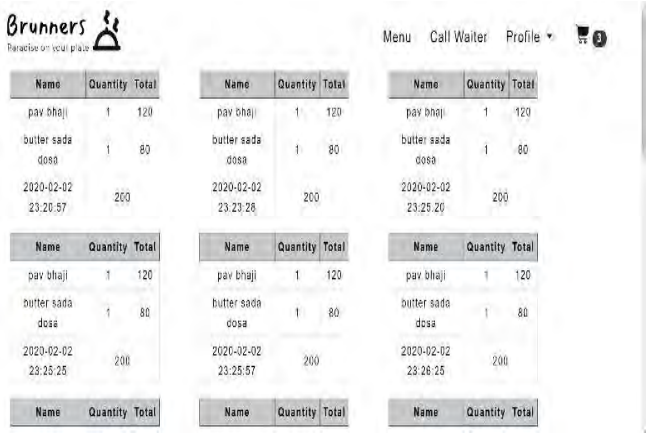


Fig.9 Previous Orders Page

This page gives details about the food items ordered by the customers in the past.

Fig.10 AR Snapshot

The above snapshot is of food item which is to be integrated with the website so when customer wishes to view the food item, it can be shown with the help of AR technology.

IX. CONCLUSION

By passage of time and popularity of the culture of using smartphones, the influence of this tool in all aspects was highlighted so that it has become an inseparable part of human life. The use of this capacity in the matters of business and directing users to the optimum and proper use of this tool is what customers and business owners are looking for. However, users tend to use technology in different situations more and more by increasing penetration of smartphones. A business can use this situation and benefit more that move before customers and prepare an appropriate condition in advance. Using application for ordering process causes to remove physical ordering and payment. In addition to more attraction, it provides more facilities to customers and business managers. This plan is not only to eliminate traditional ordering, but also to apply this smart ordering beside traditional method and attract customers and improve their sense of loyalty. Also, wasting time and energy of staffs will be reduced using this method and consequently customers are more satisfied.

- It is expected from the restaurant to provide Wi-Fi network to load the site. The developer will design such a system that would create two profiles, one for the manager/staff and one for the customer. The payment gateway should be well taken care of along with the security of the sensitive information.
- The project will be a new step towards digitalization in food business sector. Customers can track their order and estimate time by which they will receive order. A display has to be made available inside the kitchen for chefs to see the order.
- A ‘call’ button is also introduced in updated versions, which will notify waiter to visit a table. The motto of the project to optimize waiting system at restaurants is very well achieved. In future can also be helpful for food courts at some mall, where there exist multiple outlets.

REFERENCES

[1] Kunal Raut, Priyanka Khare , Aishwarya Kamble, Sachin Deshpande. Augmented Reality in Restaurant. February 2018

[2] Lei Xie , Chuyu Wang, Yanling Bu, Jianqiang Sun, Qingliang Cai, Jie Wu , Sanglu Lu. TaggedAR: An RFID-Based Approach for Recognition of Multiple Tagged Objects in Augmented Reality Systems. May 2019

[3] Mr. Aditya R Mr. Anirudh K.R Mr.Mahinder P Ms. Deepa S.R An Implementation of NFC based restaurant table service May 2017

[4] Ms. Farhah Saidin Noor Dayana Abd Halim & Noraffandy Yahaya. A review of research on Augmented reality in Education : Advantages and Applications. March 2015

[5] Caudell, T. and Mizell, D. “Augmented reality: An application of heads-up display technologyto manual manufacturing processes”,In proceedings of the Hawaii International Conferenceon System Sciences, volume 25, pages 659-659. IEEE INSTITUTE OF ELECTRICAL NDELECTRONICS, 1992.

[6] Chen, D. et al. “Streaming mobile augmented reality on mobile phones”, In Proceedings of the 2009, 8th IEEE International Symposium on Mixed and Augmented Reality, pages 181-182. Citeseer,2009.

[7] Gunnarsson, A. et al. “Visualization of sensor data using mobile phone augmented reality”, In Proceedings of the 2006 Fifth IEEE and

[8] ACM International Symposium on Mixed and AugmentedReality (ISMAR'06)-Volume 00, pages 233-234. IEEE Computer Society, 2006.

[9] Hong, Y. et al. “Mobile pointing and input system using active marker”, In IEEE/ACM InternationalSymposium on Mixed and Augmented Reality, 2006. ISMAR 2006, pages 237-238. 2006.

[10] Papagiannakis, G., Singh, G. and Magnenat-Thalmann, N. “A survey

[11] of mobile and wirelesstechnologies for augmented reality systems”, Computer Animation and Virtual Worlds, 19(1):3,2008.

[12] Seichter, H. et al. “Multitouch interaction for Tangible User Interfaces”, In Mixed and AugmentedReality, 2009. ISMAR 2009. 8th IEEE International Symposium on, pages 213-214. 2009.

[13] Ashwini J, Ayisha Shetty, Chaithra, Lasya K Rao Asst. Prof. Ramya D Shetty.” Automated Food Ordering System”. April-2018

[14] Adithya R, Abhishek Singh, Salma Pathan, Vaishnav Kanade. “
Online Food Ordering System” December 2017

E-Voting Using Blockchain

Ritika Arya, Neha Rathod, Kajal Rathod
Thakur College of Engineering and Technology Mumbai, India
aryaritika2101@gmail.com, nehakrathod48@gmail.com, kajalrathod086@gmail.com

Abstract- **Building an electronic democratic framework that fulfills the lawful necessities of officials has been a test for quite a while. Conveyed record advances is an energizing mechanical headway in the data innovation world. Blockchain advances offer a limitless scope of utilizations profiting by sharing economies. Security of advanced democratic is consistently the greatest concern when considering to execute a computerized democratic framework. With suchamazing choices in question, there can be no uncertainty about the framework's capacity to verify information and shield against potential assaults. One way the security issues can be possibly understood is through the innovation of blockchains. Blockchain innovation begins from the basic structural plan of the digital currency bitcoin. It is a type of dispersed database where records appear as exchanges, a square is an assortment of these exchanges. With the utilization of blockchains a safe and vigorous framework for advanced democratic can be formulated.The point of this report is to diagram our proposition to explaining the issues of computerized casting a ballot by utilizing blockchain innovation. The principle angle behind building up a blockchain based framework for casting a ballot is to give a simple to utilize methods for leading decisions for different sorts of gatherings and associations. Vote checking is likewise made simple by the application since it is simply an issue of questioning the database. A decent application that can anticipate framework disappointments on a broadly utilized stage can pick up acknowledgment among the potential clients.**

Keywords— Blockchain, Ethereum, Solidity, Hash Algorithm

I. INTRODUCTION

Casting a ballot, regardless of whether conventional expressive dance based or electronic democratic (e-casting a ballot), is the thing that cutting edge majority rules systems are based upon. In each majority rules system, the security of a political decision involves national security. The PC security field has for 10 years contemplated the potential outcomes of electronic democratic frameworks. With the objective of limiting the expense of having a national political race, while satisfying and expanding the security states of a political decision.

From the beginning of justly choosing up-and-comers, the democratic framework has been founded on pen and paper. Supplanting the conventional pen and paper conspire with another political decision framework is basic to restrict misrepresentation and having the democratic procedure recognizable and evident. Electronic casting a ballot machines have been seen as imperfect, by the security network, essentially dependent on physical security concerns. Enter blockchain innovation. A blockchain is a disseminated, permanent, indisputable, open record. These innovative highlights work through cutting edge cryptography, giving a security level equivalent as well as more prominent than any recently known database. The blockchain innovation is in this manner considered by many, including us, to be the perfect apparatus, to be utilized to make the new present day equitable democratic procedure.

In this manner, in the event that we actualize blockchain innovation into evoting framework (Blockchain-empowered e-casting a ballot (BEV)), every one of the votes can be recorded, overseen, tallied, checked, confirmed by the voters themselves and in any event, securing the voter character and protection.

Building an electronic democratic framework that fulfills the legitimate prerequisites of administrators has been a test for quite a while. Appropriated record advances is an energizing innovative progression in the data innovation world. Blockchain advancements offer an endless scope of utilizations profiting by sharing economies. This undertaking expects to assess the use of blockchain as administration to actualize dispersed electronic democratic frameworks. The undertaking elicitates the prerequisites of building electronic democratic frameworks and distinguishes the legitimate and innovative impediments of utilizing blockchain as an assistance for acknowledging such frameworks.

Security of computerized casting a ballot is consistently the greatest concern when considering to actualize an advanced democratic framework. With such grand choices in question, there can be no uncertainty about the framework's capacity to verify information and guard against potential assaults. One way the security issues can be conceivably unraveled is through the innovation of blockchain.

II. PROBLEM DEFINITION

To propose conceivably another e-casting a ballot convention that uses the blockchain as a straightforward polling station. The target of such a plan is given a decentralized design to run and bolster a democratic plan that is open, reasonable and freely obvious. Web-based

democratic - like a choice to paper polling forms or electronic democratic machines - has been recommended as an approach to support the number of dynamic voters, however perhaps even address political race security and respectability issues.

III. PROPOSED SYSTEM

Voter: A voter, distinguished by one's open key, Vipub, is viewed as a substance that is allowed to make a choice towards one of the up-and-comers. The voter, during the political race, will be called to settle on a decision, Ci that can run from a lot of predefined decisions, to fight messages

Central Authority (CA): All together for the e-casting a ballot convention to give confirmation that lone qualified voters can cast a ballot, it was considered essential for a Focal Power to be presentedFor a client to be made a decision about qualified, one must validate oneself to the Focal Position, CA, and get a token that demonstrates one's qualification to cast a ballot. Vote: A vote is a message of predefined structure which will be compared to a bitcoin exchange. Ethereum stage It takes the vote of the client and gives updation-in the democratic mean a specific up-and-comer.

The development incorporate stages like: During the instatement of a political race a beginning agreement must be put on the blockchain. This beginning agreement contains the entirety of the data that will be important to approve the squares as they are put, guaranteeing that none are put after time, that none are put without the proper marked token.Voting stage: It is where the voter claims for a vote. At the point when they make their accommodation, the voter must incorporate the segment portions of this marked token (division of the token into its segment parts is too expensive gas on the blockchain to be viewed as possible).

Checking stage: When the political decision has finished up, votes should be tallied. This stage enables every client to make just a single choice. Logging off: The enlisted client can log out from the framework.

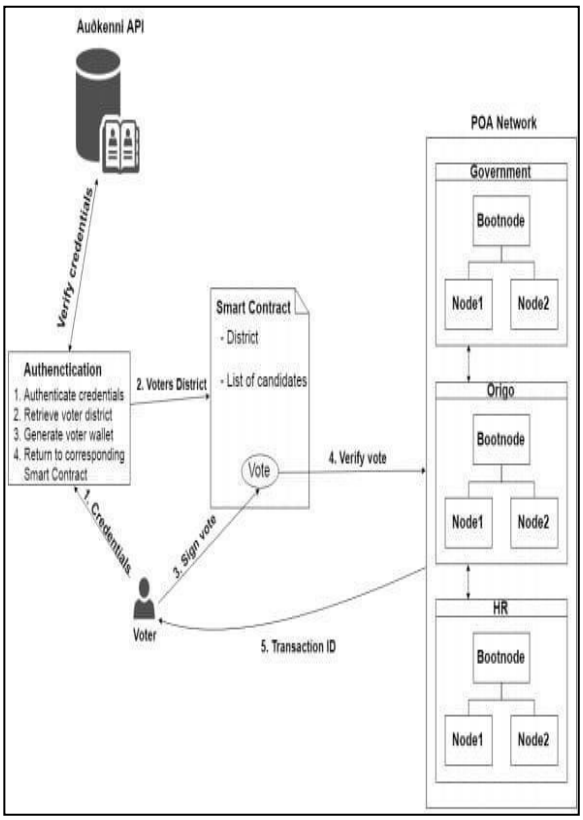


Fig 1: Voter Authenticates Himself And Casts Vote
Image References:[5]

IV. METHODOLOGY

Agile programming improvement alludes to a gathering of programming advancement systems dependent on iterative advancement, where necessities and arrangements develop through cooperation between self-sorting out cross-functional groups. Deft strategies or Coordinated forms for the most part advance a taught venture the executives procedure that supports visit assessment and adjustment, an authority reasoning that empowers cooperation, self-association and responsibility, a lot of building best rehearses proposed to take into account quick conveyance of top-notch programming, and a business approach that adjusts improvement to client needs and friends objectives. Nimble advancement alludes to any improvement procedure that is lined up with the ideas of the Lithe Declaration. The Statement was created by a gathering fourteen driving figures in the product business, and mirrors their experience of what approaches do and don't work for programming advancement.

Scrum is a subset of Agile. It is a lightweight procedure structure for light-footed improvement, and the most generally utilized one. A "process system" is a specific arrangement of practices that must be followed all together for a procedure to be predictable with the structure. (For instance, the Scrum procedure system requires the utilization of improvement cycles called Dashes, the XP structure requires pair programming, etc.)

A Scrum procedure is recognized from other coordinated forms by explicit ideas and practices, isolated into the three classes of Jobs, Ancient rarities, and Time Boxes. These and different terms utilized in Scrum are characterized beneath. Scrum is regularly used to oversee complex programming and item advancement, utilizing iterative and gradual practices. Scrum essentially expands profitability and diminishes time to benefits comparative with exemplary "cascade" forms. Scrum forms empower associations to alter easily to

quickly evolving prerequisites, and produce an item that meets advancing business objectives. A light-footed Scrum process benefits the association by pushing it to

- Increase the nature of the expectations.
- Cope better with change (and anticipate the changes)
- Provide better gauges while investing less energy making them.
- Be more responsible for the task calendar and state.

V. FLOWCHART

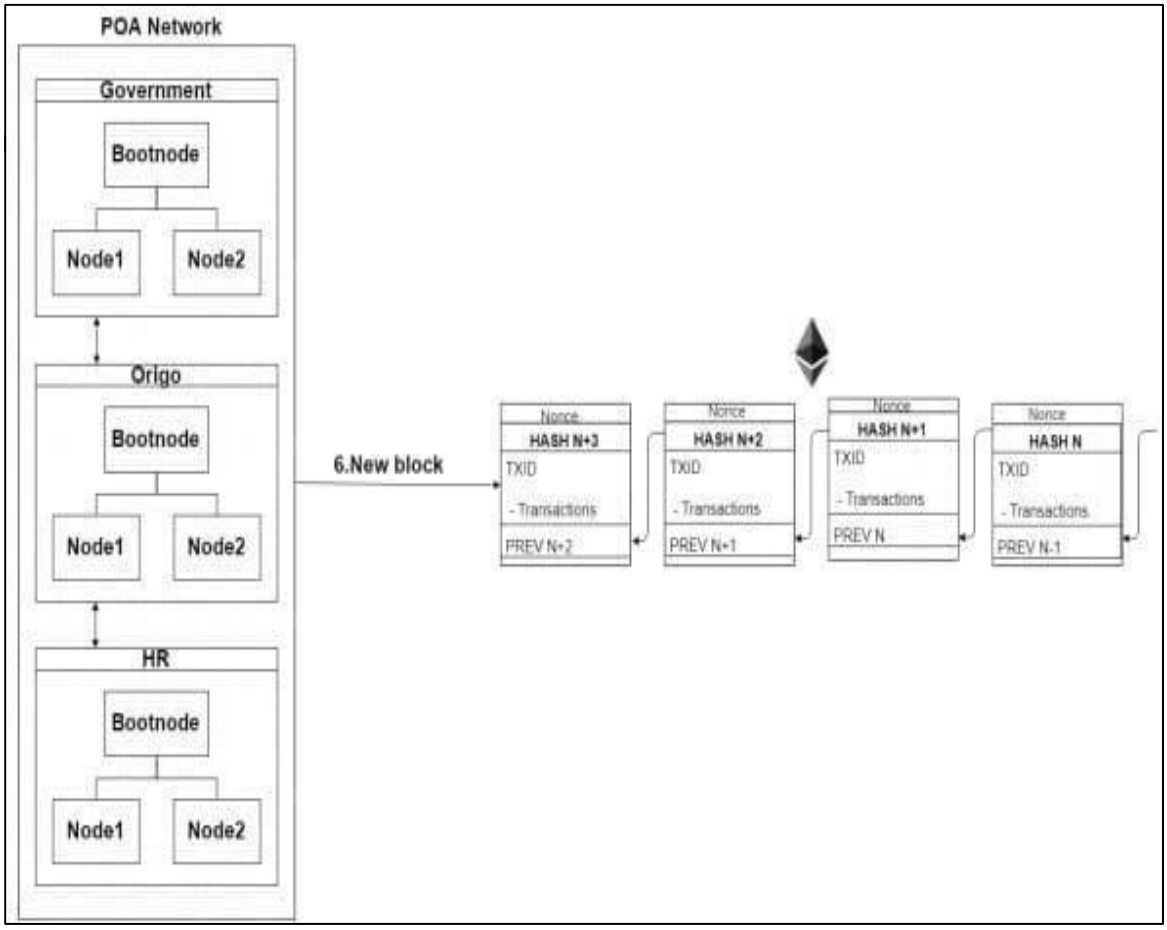


Fig 2: Block Added To Blockchain
Image References: [6]

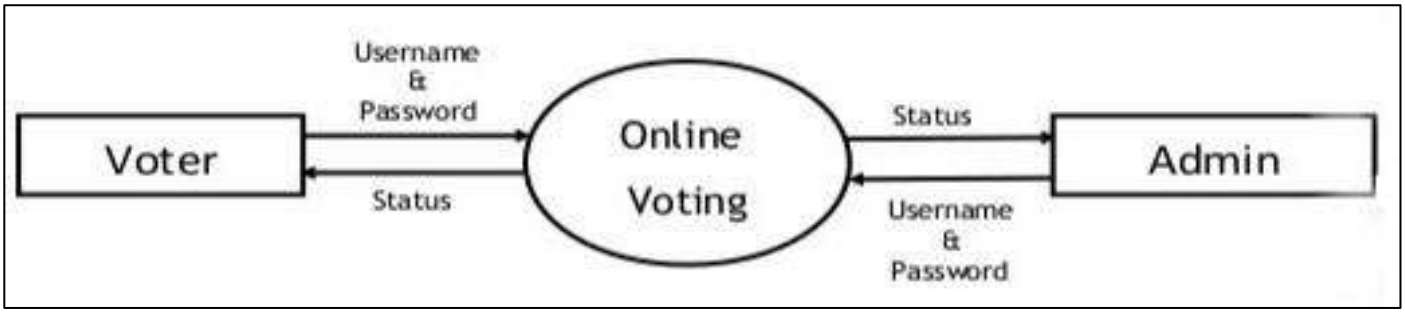


Fig 3: DFD Level 1

VI. RESULTS

Fig 4: Home Page For Casting Vote

In the above image, the home page is displayed where initially no vote is casted. The user will login and then cast a vote for any contestant i.e. Tom & Jerry.

Fig 5: First User's Vote Casted

After the user login, the vote has been casted to the contestant Tom. Henceforth, no vote can be casted to any of the other contestants.

Fig 6: Vote Casted By Another User

Another user logged in and casted a vote to the contestant Jerry. Once the vote is casted, no vote can be casted to any of the other contestants. If the user tries to vote again from the same account then an error message will be displayed.

VII. FUTURE SCOPE

Our election scheme allows individual voters to vote at a voting district of their choosing while guaranteeing that each individual voters vote is counted from the correct district, which could potentially increase voter turnout.

The future scope of the project includes the improvement in the security level of the complete system. In addition to that it would be necessary to meet some other confidential primitives to improve the security level of online voting system. The future improvement can also be done for voice verification, system crash and power failure so that the voters can vote with more assurance.

We will introduce aadhar card based voting in which a person can vote only by using his aadhar card number which is unique for everyone and by which one voter can vote only once.

VIII. CONCLUSION

Our implementation was tested on the Ethereum testnet network with different types and sizes of ballots. We used the smart contracts in Ethereum blockchain to keep a record of every user in our system as well as all the ballots and the information regarding them. We also utilized the smart contracts to achieve access control.

Adapting computerized casting ballot frameworks to make the open discretionary procedure less expensive, quicker and simpler, is a convincing one in current society. Making the discretionary procedure modest and speedy standardizes it according to the voters, expels a specific power hindrance between the voter and the chosen authority and puts a specific measure of weight on the chosen authority. It additionally opens the entryway for a more straightforward type of majority rules system, enabling voters to express their will on singular bills and suggestions.

Utilizing an Ethereum private blockchain, it is conceivable to send several exchanges for every second onto the blockchain, using each part of the shrewd agreement to facilitate the heap on the blockchain.

We have proposed an electronic democratic framework dependent on the Blockchain innovation. The framework is decentralized and doesn't depend on trust. Any enrolled voter will be able to cast a ballot utilizing any gadget associated with the Web. The Blockchain will be openly undeniable and conveyed in a way that nobody will have the option to degenerate it.

IX. ACKNOWLEDGMENT

We sincerely thank our guide Mr. Aaditya Desai for his guidance and constant support and also for the stick to our backs. We also thank the project coordinators for

arranging the necessary facilities to carry out the project work.
We thank the HOD, Dr. Rajesh Bansode, the Principal, Dr.B. K. Mishra, and the college management for their support.

REFERENCES

[1] Aanjana Devi.S, Dr.Palanisamy.V and Anandha Jothi.R, Confidential E-Voting System Using Face Detection and Recognition, Department of Computer Applications, Alagappa University, Karaikudi, International Journalof Engineering and Techniques - Volume 3 Issue 4, July-Aug 2017

[2] Ms. Anagha Hegde, Ms. Chetana Anand, Ms. Jyothi B, Mobile Voting System .UG Student, Department of Information Science and Engineering, Dr. Ambedkar Institute of Technology, Bangalore, India International Journal of Science, Engineering and Technology Research (IJSETR) Volume 6, Issue 4, April 2017, ISSN: 2278 -7798

[3] Trisha Patel, Maitri Chokshi, Nikhil Shah , Smart Device Based Election Voting System Endorsed through Face Recognition., 2014 Department of Computer Science & Applications,Charotar University of Science & Technology, India, International Journal of Advanced Research in Computer Science and Software Engineering,
Volume 3, Issue 11, November 2013 ISSN: 2277 128X

[4] Friðrik Þ. Hjálmarsson, Gunnlaugur K., Blockchain-Based E-Voting System, Hreiðarsson School of Computer Science Reykjavik University, Iceland {fridrik14, gunnlaugur15}@ru.is

[5] https://www.economist.com/sitses/default/files/plymouth.pdf

[6] https://spectrum.ieee.org/telecom/internet/online-voting-isnt-as-flawed-as-you-thinkjust- ask

Natural Food Recognition using Deep Neural Network

Ritika C Enagandula, Shipra S Guin, Simran A Mehta, Mr. Namdeo Badhe
Thakur College of Engineering and Technology, Kandivali (East), Mumbai, India – 400101.
ritikaenagandula28@gmail.com, shipraguin@gmail.com, mehtasimran234@gmail.com, namdeob.badhe@thakureducation.org

***Abstract*— The process of identifying food item from an image is quite an interesting field with several applications. Since Food surveillance plays a significant role in health-related problems, it is becoming more necessary in our day-to-day life. In this paper, the main aim of the system is to classify the image of the food item using convolutional neural networks. We have implemented three specific algorithms – Resnet, SequentialNet and Google Inception V3. The accuracy of the Google Inception V3 is more efficient as compared to the other algorithms. The dataset used is – 'Food-101' to scan with the natural food item, thereby getting the most efficient trained algorithm for the system. Accordingly, recognition will be performed on each sectioned item utilizing deep learning models that are trained based on huge food datasets. Lastly as the future scope, volume and weight estimation can be performed on each recognized food item, followed by the nutrient assessment.**

***Index Terms* — Convolutional neural networks; Food 101 dataset; Food classification; Food Image recognition; ResNet; SequentialNet; Google Inception V3.**

I. INTRODUCTION

In the current age, people are cognizant about their nourishment and diet to keep away from either up and coming or existing diseases. Since people are subject to advance technologies, arrangement of an application to consequently monitor the people's food consumption, helps in numerous perspectives.[1] It builds the familiarity with individuals in their nourishment propensities and diet. In the course of the recent two decades, explore has been centered around naturally perceiving the nutritional information and their healthful data from pictures caught utilizing computer vision and deep learning[2]. Image recognition is among the most important fields of image processing and computer vision. Food image classification is a distinct branch of image recognition problem[1]. In recent times people are more conscious about their health. A majority share of the people is indulged in overeating and not being dynamic enough. Given how occupied and focused people are today, it's easy to neglect and monitor the food that they eat. This only expands the significance of legitimate arrangement of nutrients depending upon the proper natural food image classification using advanced algorithms.

With the speedy development of our society, additional attention has been paid to the standard of life, particularly the type of food that people tend to eat[3]. However, classifying food manually isn't applicable to the present fast-tempo society any longer. An automatic food organization with enlarged accuracy, improved speed and reduced cost is desperately required. In recent years, laptop vision systems are used

immensely in food recognition strategies.

Therefore, the aim is to develop such a system that will help to fulfil all these needs. The project mainly aims at recognizing the natural food image with an efficient resolution using various algorithms of high accuracy. A system which can classify food from image is necessary for a dietary assessment system. Classification of food images is extremely challenging since the dataset of food images is non-linear. This paper proposes a method that can classify food types with images. This project proposes the use of convolutional neural network to classify food images. The CNNs are an effective class of neural networks that is highly effective at the task of image classifying, object detection and other computer vision problems. We classified a food dataset named 'Food101' consisting different food categories. Later part of the image recognition system which will be useful for estimating the nutrients and calories which is of significance in order to have a daily track of the food intake. This is important as most of the teenagers eat a large amount of junk food which deteriorates their health due to which there may be high risk to their lives leading to Heart attack at a very small age. According to medical studies, overweight and obesity are generally the result of energy imbalance between the number of calories burned by the body for daily life activities and the amount of energy coming from consumed foods. The clinical treatment and prevention of overweight and obesity require the patients to measure, record, and control daily food intake. All of this can be avoided by recording everyday meals.

Hence food recognition is very important for classifying the food image and thereby help in the health control. Smartphone based food logging will done using this project where in the food item that a person eats is first captured using camera of the smartphone to showcase the classified image and recognize the food item. The research on food classification is focused around growing constant applications which catch pictures and train the AI models immediately. However, the efficient extraction of information from the images of food still remains a challenging issue. In this paper, an attempt has been made to classify the input images of food into different food categories with the help of

convolutional neural networks (CNNs)[5]. Since the CNNs can handle a large amount of data and can evaluate the features consequently, it has been used for the task of image classification of the food. The Food 101 dataset has been chosen as the working database for this methodology.

The task of the food detection system is first initiated with packaged food items. Recently, CNNs have been utilized to distinguish edges and shapes with extraordinary achievement. Here, motivated by this achievement the strategy is to take different food items by utilizing a CNN that distinguishes food outskirts[9].

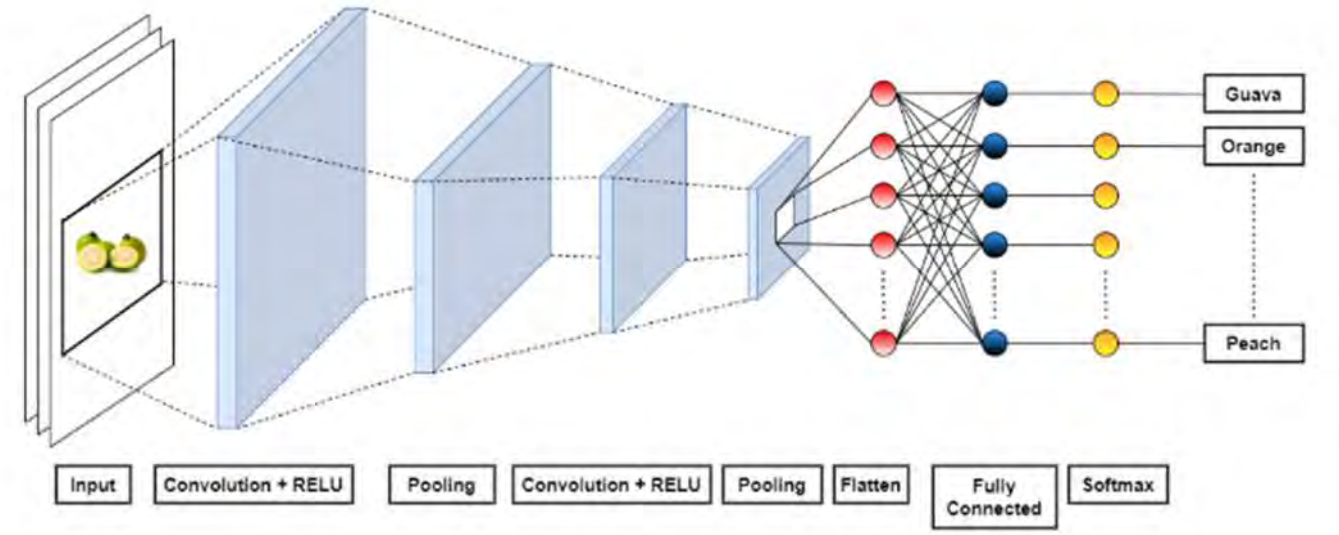


Figure 1: Convolutional Neural Network Layers for the proposed system

II. Related Work

To ease the weaknesses of these clinical strategies, researchers have been trying to come up with improved systems. A portion of these techniques requires the individual to take a photo of the food in order to recognize the food item using the natural image using the pre-trained network model to get the maximum accuracy[4]. Be that as it may, arranging food physically isn't pertinent to the present fast-tempo society anymore. An automatic food organization with expanded precision, improved speed and diminished cost is desperately required. In recent years, laptop vision systems are used immensely in food recognition strategies.

Therefore, we plan to grow such a system, that will satisfy every one of these necessities. The recognition is significant to This is significant as most of the young people eat a lot of low-quality food which deteriorates their wellbeing and leads them to diseases like Blood pressure, Obesity, Diabetes and so forth. Because of which there might be high risk to their lives prompting heart attack at a very young age. Medical studies uncovered that overweight and obesity are commonly the aftereffect of vitality lop-sidedness between the calorie originating from the consumed food and the quantity of calories consumed by the body for day by day life exercises. The clinical treatment and counteractive action of overweight and weight require the patients to gauge, record, and control day by day food consumption. The track of this can be maintained by recording each day meal.

Despite the fact that perceiving food among many classes is an exceptionally complex issue, figuring out how to identify their fringes is frequently simpler and less subject to the thought about classes. Along these images, the technique can get the remarkable execution of CNNs and always improve through extra preparing without depending on the food recognition result.

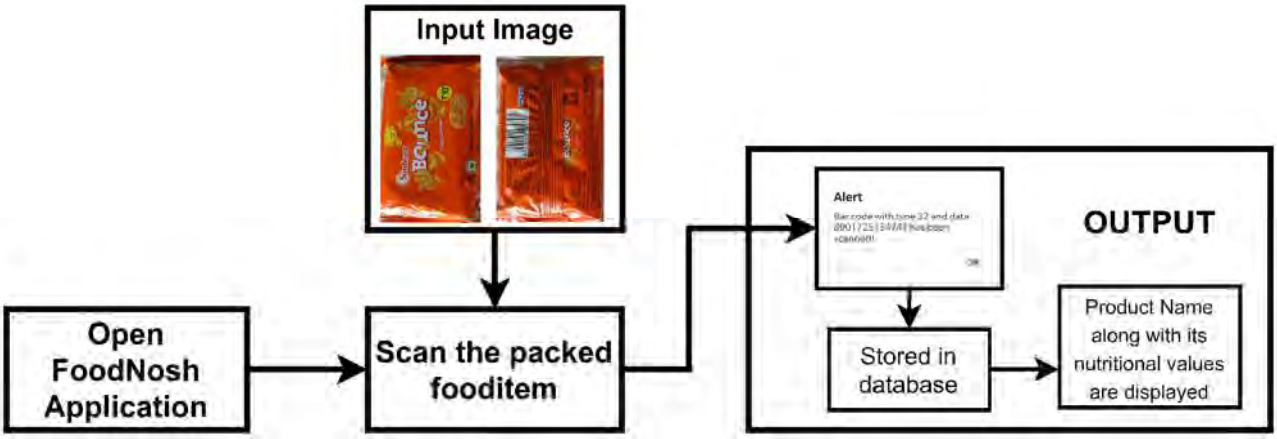
The present work intends to join a portion of the above strategies together, that makes a food classification system, that predicts the class of food the picture is in and recognizes the same. In the past few years, a variety of methods has been suggested for image classification and identification. CNN is used to automatically learn and extract features from the images which are then classified. However, the training process is very time-consuming, even using GPUs[9].

III. Literature Survey

Categorical Image Classification needs thousands of images to train the model for classification of food items [1]. Various types of food with different color and texture reflect the fact that the food image recognition is considered a challenging task. Also, same food items in different regions might have an all together a different look which makes it difficult to identify the food item. This crucial task is done using deep learning. However, deep learning has been widely used as an efficient image recognition method, and CNN is the contemporary approach for deep learning to be implemented [3].

CNN's really use next to no pre-processing. This implies that they might take in the filters that there must be a chance to be hand-made in other calculations.

CNN's basic tasks are feature recognition, picture classification, and recommendation frameworks. A recommendation system can be defined as a system that will recommend items to the users/customer within an environment/system depending on their past activities [10].



User needs to take the picture of the food using their mobile devices before and after the meal for the precise estimation of calories. The subsequent stage of the system is segmentation, each image will be broken down to separate different sections of the food image portion. Out of different tools accessible for segmentation, colour and texture segmentation tools are utilized for the successful estimation. Different food features including size, shape, shading and surface will be extracted and sent to classification step. Therefore, utilizing the above steps, segment of nourishment is recognized. At last, by evaluating the area of the food portion and utilizing some nutritional tables, the calorie estimation of the food will be extracted [5].

IV. Background

The algorithms used for the system are ResNet, SequentialNet and Google Inception V3. The main task is to recognizer the food item using the most efficient neural network model to recognize the food item and thereby estimate the nutrients. The main logic behind the recognition system is transfer learning and since according to the results Google Inception V3 is the most efficient algorithm with a greater accuracy, we use the transfer learning which uses a pre-trained neural network. Google Inception V3 model is divided into two parts a) Feature extraction with a convolutional neural network. b) classification part with fully-connected and softmax layers.

The image will be extracted and scanned using the pre-trained model to recognize the natural food image and give us the final result. Tensorflow is one of the libraries used for image classification in deep learning. Tensorflow is an open source software library which was developed by Google in the year 2015 for numerical computation. Tensorflow is specially designed to enable fast experimentation with deep neural networks[9].

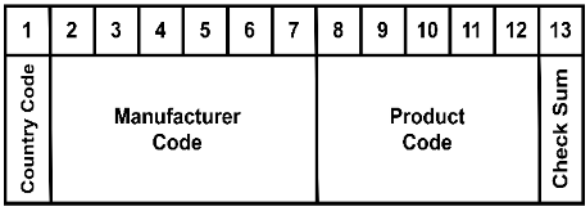


Figure 2: Barcode EAN 13 Format

Proposed System

Phase 1: For Packaged Food Items

In this particular phase we started with a proper solution for the packaged food items. Where we scan the barcode of the packaged food item and get the respective data for each product. In this system, we provide a recommendation system where it will help the user to decide whether to consume that food item or not and thereby give another option as well. The system works as mentioned below in the flowchart below. Where the product is scanned using our application "FoodNosh" after the successful login process the user will be able to scan the product and later the application will display the alert message with the details of the type of the food item and the corresponding barcode number. After the scan process the data will be stored in database in with all the details including all the nutritional values that will be extracted from the barcode. This will make it convenient to store that data of all the packaged food items in the database and thereby get a proper comparison with respect to which food item should be recommended instead of the one which is recently scanned by the user.

This will be a part of the recommendation system where the comparison will take place according to the medical data registered by the user and thereby, we can determine which food item shall be preferred.

This phase gives us a base to look after all the nutrients and compare them and thereby give a result to the user which satisfies them and their health at the same time. In this phase the nutrients and other values are

constantly maintained in order to track the total consumption of the user.

Phase 2: For Unpacked Food Items

The objective of our system is to design an automated application to classify food items running on a smart phone with working camera that makes it easy to record food item that are consumed during the day. Our objective isn't to essentially have high accuracy, because as explained such accuracy is not possible in real. Obviously, the more accurate the system is the better the final result, and this is the reason in this paper we have attempted to identify various food items precisely as possible. In any case, it is significant to understand that high accuracy is not possible while dealing with the images of the food item only.

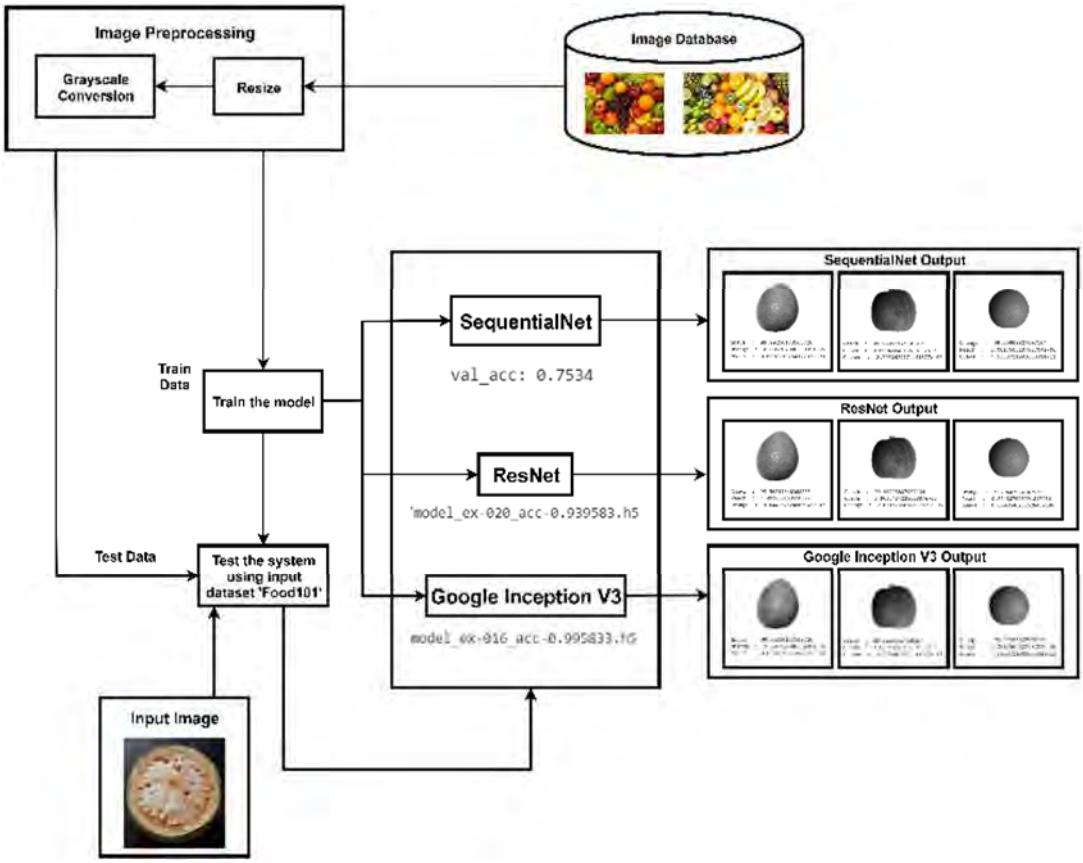
We propose a food image recognition system with various variant of CNN algorithms which can deal with different food items classification. Given an input as a food picture, first, the CNN can achieve all the recognition steps including highlight extraction, shift and distortion invariance and classification, and afterward, gives the output.[8].

Every classification system starts with the selection of the database which is in coherence with your systems requirements. For our system we chose to go with the Food-101 dataset which contains various RGB images of different kinds of food items. The images of the dataset are pre-processed before training the model for classification. For example, we perform conversion of the images to grayscale on the dataset. In addition, it's required to resize the image for their optimal use of size 70 x70. The original dataset we use is Food-101 which is an open 120-class food image dataset that is a food dataset that we establish containing only the images we want to train for and thereby increasing the accuracy and reducing the error rate and also to perform the testing on the trained model using the same dataset.

Here the main aim is to identify the most accurate algorithm to perform classification of the food items that are to be consumed. So, we have chosen to perform classification using 3 algorithms from the various options available. The choice of the first algorithm, is obvious SequentialNet because it is the most basic CNN algorithm. Since the main aim of the system is to provide the maximum accuracy, transfer learning algorithm is the most practical solution to go with. Therefore, we chose to go with ResNet and Google Inception v3 algorithm. We would apply training of the same dataset on 3 different algorithms and compare the accuracy of the same and also conclude the best amongst the 3 algorithms.

We use stochastic gradient decent to train our model with a batch size of 120 examples, initial learning rate of 0.01 and epoch of 200. A small learning rate is very important for the model to learn and it can reduce the model's training error. The classifier we use is SequentialNet because this probabilistic, linear classifier is effective in the model. We train the network with the trainset of 200 images, which takes about an hour.

Now, we are using the ImageAI library of python to implement the transfer learning algorithms of deep learning. Transfer learning algorithm is a model wherein knowledge used to solve a particular problem



is used to solve a different but related problem. ResNet and Google Inception v3 algorithms of transfer learning problem is used to classify the food item for the system[10].

Residual neural network (ResNet) is an artificial neural network (ANN) of pre trained neural network that builds on constructs obtained from pyramid cells of the cerebral cortex. It does this by skipping connections or by taking shortcuts to jump over some layers. ResNet does this using shortcut connections by directly connecting input of nth layer to some (n+x)th layer.

Google Inception v3 is an ANN of pre trained neural network used widely for image classification. The model itself is made up of symmetric and asymmetric building blocks including convolutions, average pooling, max pooling, concats, dropouts, and fully connected layers. Batchnorm is used extensively throughout the model and applied to activation units. Loss is computed using softmax.

Agile is the methodology adopted for the system, it is because of which we can adapt the change quickly to win the market, and this is a core competency. Therefore, we will use the agile methodology in order to get the result at a faster rate and thereby do not have to repeat the procedure if we come across some default errors which happens in Waterfall Model.

V. Analysis

The proper analysis for the system will help us to improve the ability to define the scope of the project and manage the project within that scope. It will also help us to learn how to identify and sequence the tasks, estimate duration of tasks, control variances, manage costs, and utilize resources. This will also help in understanding the qualitative and quantitative techniques for identifying, analysing, and mitigating risk, as well as the best ways and times to apply these techniques to the project environment[6] .

Despite the quality detection and inspection of the food item is not strictly related to the application domain of dietary food monitoring, we have decided to include information on this application domains so that the user can have a better overview of what has been done in the context of food image.

The system will give intensive instruction in project management fundamentals across the entire project life cycle. It will also provide demonstrated techniques and viable devices for planning, executing, and controlling an assortment of projects. It will also offer detailed and sophisticated instruction in the critical areas of scheduling key events, controlling costs, and managing risks.

The confusion matrix will show the accurately versus inaccurately named classes. From results, it is discovered that the CNNs are progressively suitable for image classification.

The system is meant to do a proper research of few deep learning algorithms and thereby getting the most efficient algorithm and comparatively a better accuracy. The algorithms provide features, for example, shifting and Max-pooling which give better classification rate for image classification than convolutional neural systems[7]. Convolving the image permits feature extraction, regardless of o its direction and position in the image. By completing this project, we will be able to configure a Food Detection and Nutrients Identification system to work (in a limited manner) to support project management.

We would need to fully understand the commercial situation and how to make use of SDLC, inspection and other crucial processes.

VI. Result And Discussion

The below picture shows the comparison about the accuracy and the effectiveness of the Algorithms – ResNet,
SequentialNet and Google Inception V3.

The trial method to implement the various algorithms was successful to understand which algorithm provides the best efficiency. Other algorithms like, MobileNet, Vgg19 which seem to be more efficient and will satisfy the accuracy required for the project.

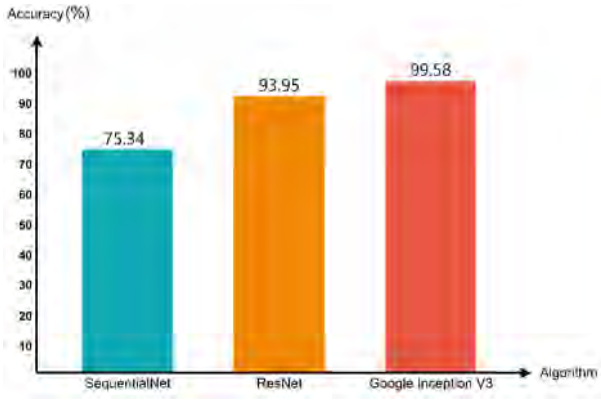


Figure 3: Comparison of the algorithms based on accuracy

The remaining work after the procedure of classification is mapping the food item names to a calorific value. This will be accomplished effectively by scratching the web for the average calories value will be considered of the food items things per unit weight[3]. The average calorie values are considered for the various classes of the food items[3].

VII. Conclusion

From the above results the transfer learning algorithms clearly improves the accuracy for the image classification. Since in transfer learning you do not need to train the classification model from scratch as it contains a pre-trained model which is not available with the basic SequentialNet 3- layer algorithm. Also, the number of layers is quite less in the SequentialNet algorithm which is a major factor resulting in less accuracy. The performance of CNNs depends heavily

on multiple hyper parameters like the number of layers, number of feature maps in each layer, the use of dropouts, batch normalization, etc. Hence training of the model mainly includes adjusting the model hyper parameters by conducting lots of experiments. Thus, this paper helps in identifying and understanding the best algorithm that can be used for food image recognition.

VIII. Future Scope

It is not always possible to implement each algorithm using transfer learning as some problems are dynamic in nature where the pre trained model would act as a disadvantage in that scenario. Here the SequentialNet would be the most appropriate choice for this type of scenario.

In this system the recommendation for the user works well until and unless there is a proper internet connectivity available. If the user lives in a rural area, the application "FoodNosh" will not work up to the expectation. Thereby, storing the data of the food items in the database will be difficult in that case of the user and this will stand as a drawback for the system. Acknowledgment

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have completed the project successfully.

We would like to thank everyone for their guidance. We sincerely thank our Principal, Dr. B. K. Mishra, Vice Principal, Dr. Deven Shah, Mentor Dean, Dr. Kamal Shah and HOD, Dr. Rajesh Bansode for always encouraging us to do our best. We are highly indebted to our guide Mr. Namdeo Badhe who supported and constantly supervised us through this project and helped us in not only completing this project but also provided us with sample amount of knowledge that was really beneficial to us. We also thank the project coordinators for arranging the necessary facilities to carry out our project work.

We are thankful to and fortunate enough to get constant encouragement, support and guide from all teaching staff of IT Department who helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in laboratory for their timely support.

We would like to express our gratitude towards our parents for their kind cooperation and encouragement which helped us in completion of this project.

Reference

[1] Mrs. Arpana Mahajan, Dr. Sanjay Chaudhary, "Categorical Image Classification Based on Representational Deep Network (RESNET)", IEEE Transactions on Instrumentation and Measurement, 2019.

[2] David J. Attokaren, Ian G. Fernandes, A. Sriram, Y.V. Srinivasa Murthy, and Shashidhar G. Koolagudi, "Food Classification from Images Using Convolutional Neural Networks", Institute of Electrical and Electronics Engineers (IEEE), November 2017.

[3] Stergios Christodoulidis, Marios Anthimopoulos and Stavroula Mougiakakou, "Food Recognition for Dietary Assessment Using Deep Convolutional Neural Networks", Springer International Publishing Switzerland, 2015.

[4] Hokuto Kagaya, Kiyoharu Aizawa and Makoto Ogawa, "Food Detection and Recognition Using Convolution Neural Network" ACM 2014.

[5] Joachim Dehais, Marios Anthimopoulos and Stavroula Mougiakakou, "Food Image Segmentation for Dietary Assessment", Association for Computing Machinery (ACM), 2016.

[6] Parisa Pouladzadeh, Shervin Shirmohammadi and Rana Almaghrabi, "Measuring Calorie and Nutrition from Food Image", IEEE Transactions on Instrumentation and Measurement, August 2014.

[7] Hafiz Muhammad Zubair Hasan, Hammad Khan, Talha Asif, Sohaib Hashmi, Muhammad Rafi," Towards a transfer learning approach to food recommendations through food images", ACM 2019.

[8] Sinno Jialin Pan and Qiang Yang," A Survey on Transfer Learning" Institute of Electrical and Electronics Engineers (IEEE), October 2010.

[9] Md Tohidul Islam,B.M Nafiz Karim Siddiquie, Saidur Rahman, Taskeed Jabid, "Image Recognition with Deep Learning",IEEE 2018.

[10] Jon Noronha, Eric Hysen, Haoqi Zhang, Krzysztof Z. Gajos, "PlateMate: Crowdsourcing Nutrition Analysis from Food Photographs", ACM Symposium on User Interface Software and Technology (*UIST*), October 2011.

SQL using NLP in Virtual Reality

Satpal Singh, Aniket Karna, Shruti Mathur
Thakur College of Engineering and Technology, Mumbai, India
satpalsingh101070@gmail.com, aniketkarna@gmail.com,

Abstract— **"SQL using Natural Language Processing & Virtual Reality" is an Artificial Intelligence based system which uses modern AI tools such as Machine Learning, Natural Language Processing, Virtual Reality ,etc. to compute and perform various tasks and activities by using Intelligent agents that accept input from user in Natural Language and perform SQL related functionalities. The system works by creating a Virtual Environment using Unity 3D in which we create Artificial Intelligence ingrained bots that can take input from user as Speech and that input can be processed by using Natural Language Processing to convert speech to digital signals. Consequently, training the bots to perform various operations and manipulations on already created Databases to execute SQL queries just by speech recognition.**

Keywords—Natural Language Processing, Virtual Reality, Sequence to Sequence Model, Artificial Intelligence, SQL, Speech to Text, Unity 3D

I. INTRODUCTION

Setting up an organization requires a lot of resources including skilled manpower, but the risk of failure of a new organization creates a lot of insecurities. Artificial Intelligence has always been considered as a tool which can help to improve the society in various aspects. Artificial Intelligence, when integrated with other powerful technology like Virtual Reality helps to create such an environment, where we can preliminarily set up a virtual organization with least resources.

The project integrates intelligent agent system along with Virtual Reality (VR) to create a virtual environment which will be capable of performing real world organizational tasks by reducing the manpower as well as creating an easy-to-communicate interface. As a purpose of demonstration, Database commands will be executed, in the real world by providing commands in Natural Language (English), to the Virtual Reality hardware.

II. BACKGROUND

A Virtual Environment that can be used to perform various tasks on Databases and Data Warehouses by using Natural Language Processing and Artificial Intelligence Technology. While thinking for the idea we found ourselves digging around the problem of lack of cheaper and safer security solutions for our homes. Hence, we planned to use the latest technologies and expertise in the best way to make a project that would try to perform various SQL tasks on Databases by a naïve user. The proposed system uses advance machine learning and Natural Language Processing for Database Operations.

III. PROBLEM DEFINITION

Developing an Virtual Environment which can be used to compute SQL queries with the help of Artificial Intelligence integrated agents by using Natural Language Processing. The system uses Seq2Seq Model, also known as Encoder-Decoder Model to convert natural language to text which is later converted to executable SQL queries. The Virtual Reality medium provides an interactive user interface through which users can interact with intelligent agents and perform required SQL operations.

IV. NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data. NLP is an interdisciplinary field concerned with the interactions between computers and human natural languages (e.g: English) — speech or text. NLP-powered softwares help us in our daily lives in various ways, for example:

- Personal assistants: Siri, Cortana, and Google Assistant.
- Auto-complete: In search engines (e.g: Google, Bing).
- Spell checking: Almost everywhere, in your browser, your IDE (e.g: Visual Studio), desktop apps (e.g: Microsoft Word).
- Machine Translation: Google Translate. [4]

Setting up an organization requires a lot of resources including skilled manpower, but the risk of failure of a new organization creates a lot of insecurities.
Artificial Intelligence has always been considered as a tool which can help to improve the society in various aspects. Artificial Intelligence, when integrated with other powerful technology like Virtual Reality helps to create such an environment, where we can preliminarily set up a virtual organization with least resources.

Such an environment will eventually avoid a lot of risks, due to organization's failure and help to handle the entire network efficiently Artificial Intelligence has emerged as the most powerful technology in today's world which is utilized in various tasks such as Natural Language Processing (NLP), chat-bots, Object Recognition, etc.
The project integrates this technology along with Virtual Reality (VR) to create a virtual environment which will be capable of perform real world organizational tasks by reducing the manpower as well as creating an easy to communicate interface.

As a purpose of demonstration, Database commands will be executed, in the real world by providing commands in Natural Language (English), to the Virtual Reality hardware. [5]

V. INTELLIGENT AGENTS

In artificial intelligence, an intelligent agent (IA) refers to an autonomous entity which acts, directing its activity towards achieving goals (i.e. it is an agent), upon an environment using observation through sensors and consequent actuators (i.e. it is intelligent). Intelligent agents may also learn or use knowledge to achieve their goals. They may be very simple or very complex. A reflex machine, such as a thermostat, is considered an example of an intelligent agent. [6]

VI. ENCODER-DECODER MODEL

The encoder-decoder model is a way of using recurrent neural networks for sequence-to-sequence prediction problems. It was initially developed for machine translation problems, although it has proven successful at related sequence-to-sequence prediction problems such as text summarization and question answering. The approach involves two recurrent neural networks, one to encode the input sequence, called the encoder, and a second to decode the encoded input sequence into the target sequence called the decoder.

Following are some of the application of sequence to sequence models-

- Chatbots
- Machine Translation
- Text summary
- Image captioning

VII. VIRTUAL REALITY

Virtual Reality (VR) is the use of computer technology to create a simulated environment. Unlike traditional user interfaces, VR places the user inside an experience. Instead of viewing a screen in front of them, users are immersed and able to interact with 3D worlds. By simulating as many senses as possible, such as vision, hearing, touch, even smell, the computer is transformed into a gatekeeper to this artificial world. The only limits to near-real VR experiences are the availability of content and cheap computing power.

VIII. OBJECTIVE

The system works by creating a Virtual Environment using Unity 3D in which we create intelligent agents that can take input from user as Speech and that input can be processed through an Encoder-Decoder Model to convert speech to digital signals. Then, training the agents to perform various operations and manipulations on already created Databases to execute SQL queries just by speech recognition.

IX. LITERATURE SURVEY

Numbers of researches have been carried out as it has become very influential topic and has been gaining heights of increasing interest. Each paper used different method and algorithm to get good accuracy on SQL queries, Database Management, Reinforcement Learning, Virtual Reality and Artificial Intelligence.

REF NO.	PAPER TITLE	AUTHOR	YEAR	KEY FINDINGS	RESEARCH GAP
1	Natural Language Processing In VR Training Environments	Curry I. Guinn & R. Jorge Montoya	2014	It demonstrates the integration of spoken human-machine dialogue with visual virtual reality in implementing intelligent assistant and training systems. Increases retention.	Fusion of virtual reality, expert systems, and natural language processing may sometimes lead to errors.
2	VR for education and Workforce training	Daniel W. Carruth	2013	Measures user performance and user's experience. Completion time and accuracy can be easily recorded in a virtual environment at a level of detail and objectivity that is very difficult to achieve in the real-world.	Simulator sickness is a common issue with virtual reality. It is common to have dropout rates of 10% to 20% and we have observed dropout rates as high as 50%.
3	Creating Real Time Intelligent Interactive Virtual Environment	Weibin Liu, Liang Zhou & Baozong Yan	2010	Generates terrain from height map and smoothens through filtering.	Integration of autonomous virtual humans and virtual environment may lead to errors.

4	Generating Structured Queries from Natural Language using Reinforcement Learning	Victor Zhong, Caiming Xiong & Richard Socher	2016	Introduction to WikiSQL. Reinforcement learning generates higher quality WHERE clause. Also incorporates structures that reduces invalid queries. Improves execution and logical accuracy.	If the former query is provided as the ground truth, using cross entropy loss to supervise the generation would then wrongly penalize the latter query.
5	Application of Artificial Intelligence in Machine Learning	Sumit Das, Aritra Dey, Aakash Pal & Nabamita Roy	2015	Automatic Indexing for Boolean Information Retrieval Systems. Semantic Annotation of Ubiquitous Learning Environments.	Requires skill and practised labours for this task
6	Agent Supported Serious Game Environment	Theodouli Terzidou,	2016	This research study proposed and evaluated an agent supported serious game environment for a 3D collaborative game in the OpenSim platform involving university students. With regards to students' attitudes, results revealed that students provided with the agent supported game environment did not change their attitude towards the course and games.	Agent's representation: the pedagogical game agent was chosen to interact with the students in a conversational mode, without any embodied representation in order to reduce the cognitive overload and to operate in a discreet manner.

Fig: Literature Review Table

X. METHODOLOGY

Artificial Intelligence, when integrated with other powerful technology like Virtual Reality helps to create such an environment, where we can preliminarily set up a virtual organization with least resources. Such an environment will eventually avoid a lot of risks, due to organization's failure and help to handle the entire network efficiently. Artificial Intelligence has emerged as the most powerful technology in today's world which is utilized in various tasks such as Natural Language Processing (NLP), chat bots, Object Recognition, etc.

The project integrates this technology along with Virtual Reality (VR) to create a virtual environment which will be capable of perform real world organizational tasks by reducing the manpower as well as creating an easy to communicate interface powerful technology like Virtual Reality helps to create such an environment, where we can preliminarily set up a virtual organization with least resources. Such an environment will eventually avoid a lot of risks, due to organization's failure and help to handle the entire network efficiently. Artificial Intelligence has emerged as the most powerful technology in today's world which is utilized in various tasks such as Natural Language Processing (NLP), chat bots, Object Recognition, etc. The project integrates this technology along with Virtual Reality (VR) to create a virtual environment which will be capable of perform real world organizational tasks by reducing the manpower as well as creating an easy to environment will eventually avoid a lot of risks, due to organization's failure and help to handle the entire network efficiently. Artificial Intelligence has emerged as the most powerful technology in today's world which is utilized in various tasks such as Natural Language Processing (NLP), chat bots, Object Recognition, etc. [1]

Seq2seq takes as input a sequence of words (sentence or sentences) and generates an output sequence of words. It does so by use of the recurrent neural network (RNN). Although the vanilla version of RNN is rarely used, its more advanced version i.e. LSTM or GRU are used. This is because RNN suffers from the problem of vanishing gradient. LSTM is used in the version proposed by Google. It develops the context of the word by taking 2 inputs at each point of time. One from the user and other from its previous output, hence the name recurrent (output goes as input).

It mainly has two components i.e *encoder* and *decoder*, and hence sometimes it is called the Encoder-Decoder Network. [3]

The system contains agent bots that are trained by using Reinforcement Learning by using sequence to sequence algorithm and hence they learn through experience by providing them training data and then verifying the operations performed by them and rewarding the system with 0/1 reward based on the validity of the result and hence their performance increases by experience. [2]

Encoder: It uses deep neural network layers and converts the input words to corresponding hidden vectors. Each vector represents the current word and the context of the word.

Decoder: It is similar to the encoder. It takes as input the hidden vector generated by encoder, its own hidden states and current word to produce the next hidden vector and finally predict the next word.

Technologies used are:

- SQL for Database Management
- Unity for creation of Virtual Environment
- Reinforcement Learning by using Seq2Seq
- Natural Language Processing
- Python for Machine Learning

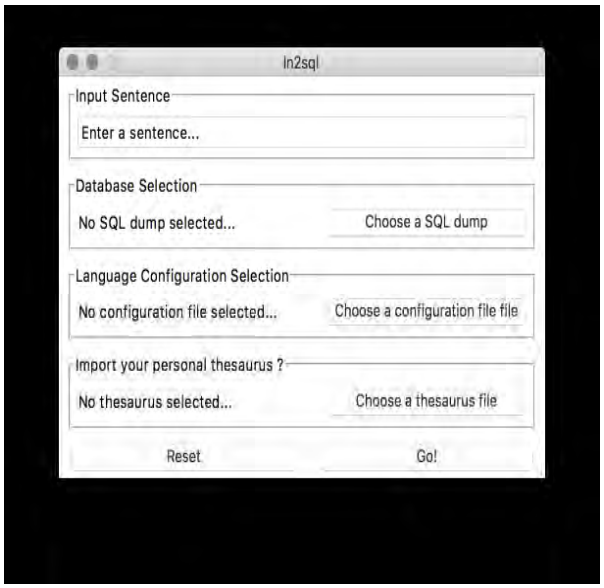


Fig: GUI Screenshot

Fig: Data Flow Diagram

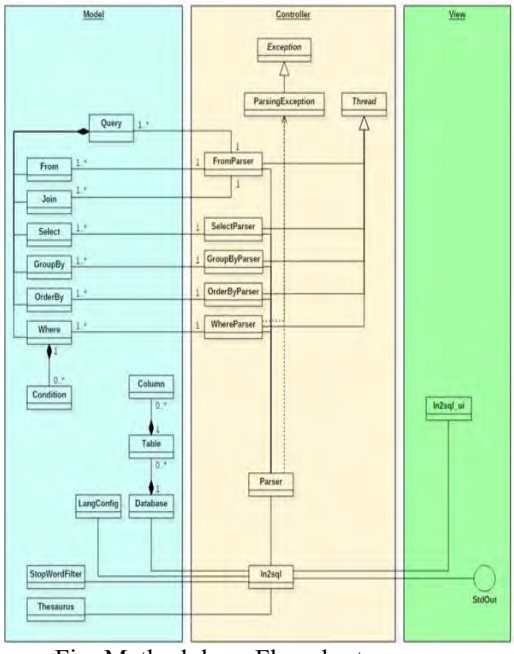


Fig: Methodology Flowchart

XI. RESULT

Developed an intelligent agent-driven system in Virtual Reality that executes SQL queries derived from Encoder-Decoder Model. This model converts Natural Language Commands into valid and executable SQL queries.

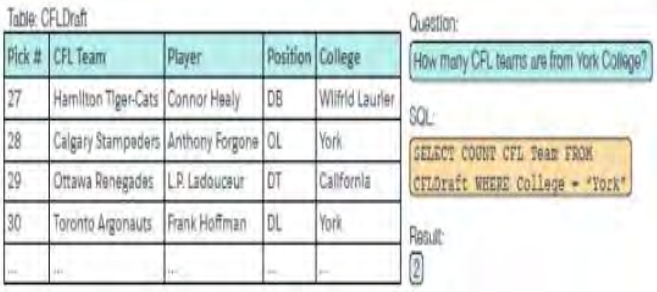


Table: CFLDraft

Pick #	CFL Team	Player	Position	College
27	Hamilton Tiger-Cats	Connor Healy	DB	Wilfrid Laurier
28	Calgary Stampeders	Anthony Forgone	OL	York
29	Ottawa Renegades	L.P. Ladouceur	DT	California
30	Toronto Argonauts	Frank Hoffman	DL	York
...	...	...	...	...

Question:
How many CFL teams are from York College?

SQL:
```
SELECT COUNT CFL Team FROM
CFLDraft WHERE College = "York"
```

Result:
2

Fig: Result Table

XII. CONCLUSION

The system works by creating a Virtual Environment using Unity in which we create bots that can take input from user as Speech and that input can be processed by using Natural Language Processing to convert speech to digital signals and then training the bots to perform various operations and manipulations on already created Databases to execute SQL queries just by speech recognition.

The system can help a user who has no previous knowledge about SQL and perform SQL queries on Database just by using speech-as-an-input methodology. This environment can help to reduce the load of tasks like Data entry, Data Manipulation in Corporate Offices and Educational institutes.

It works by creating a Virtual Environment using Unity in which we create bots that can take input from user as Speech and that input can be processed by using Natural Language Processing to convert speech to digital signals and then training the bots to perform various operations and manipulations on already created Databases to execute SQL queries just by speech recognition.

Our current system is capable of performing SQL queries on Databases by using Natural Language Processing in Virtual Reality and Artificial Intelligence.

System can be developed into an Interactive Virtual Environment that can be used to perform different tasks and activities and reflect the changes into the real world.

A Product Backlog is rarely finished. The soonest improvement of it spreads out the at first known and best-got prerequisites. The Product Backlog develops as the item and nature wherein it

will be utilized advances. The Product Backlog is dynamic; it continually changes to recognize what the item should be proper, aggressive, and helpful. On the off chance that an item exists, its Product Backlog additionally exists.

Item Backlog refinement is the demonstration of including point of interest, gauges, and request to things in the Product Backlog. This is a continuous procedure where the Product Owner and the Development Team work together on the subtleties of Product Backlog things. During Product Backlog refinement, things are audited and reexamined.

XIII. FUTURE SCOPE

Future Scope is to make a highly interactive Artificial Intelligence 3D Virtual Environment that can be used to achieve much more complicated tasks and creating more intelligent AI bots for computing and processing much more intricate data. This would help to complete demanding tasks by using Artificial Intelligence.

This system can be implemented for NoSQL, MongoDB. A module for implementing text auto-completion using Elastic-Search. We will be using solar as the searching algorithm which will counter problems such as identification of homo-sonic words, text recommendations and quick results.

The Future of Natural Language Processing explains that in NLP, machines are taught to read and interpret text as humans do. NLP is recognized as the "enabler of text analysis and speech-recognition applications." This human capability for interpreting text comes in handy for analyzing large volumes of text data. NLP is set to capture the voice of the customer. As an example of this use of NLP, think of Google Drive, where users can search documents via conversational input. With the exponential growth of multi-channel data like social or mobile data, businesses need solid technologies in place to assess and evaluate customer sentiments. So far, businesses have been happy analyzing customer actions, but in the current competitive climate, that type of customer analytics is outdated. Now businesses need to analyze and understand customer attitudes, preferences, and even moods – all of which come under the purview of sentiment analytics. Without NLP, business owners would be seriously handicapped in conducting even the most basic sentiment analytics.

References

[1] Curry I. Guinn and R. Jorge Montoya Research Triangle Institute Research Triangle Park, 27709 "Natural Language Processing In VR Training Environments", 12 September 2014

[2] Daniel W. Carruth, Center for Advanced Vehicular Systems, High Performance Computing Collaboratory Bagley College of Engineering, Mississippi State University, "Virtual Reality for Education and Workforce Training" 2013

[3] Weibin Liu, Liang Zhou, Xingqi Liu, Weiwei Xing, Baozong Institute of Information Science, School of Computer and Information Technology Beijing Jiaotong University "Creating Real-time Intelligent Interactive Virtual Environment" International Conference on Intelligent Computation Technology and Automation, 2010

[4] Victor Zhong, Caiming Xiong & Richard Socher, "Generating Structured Queries from Natural Language using Reinforcement Learning" 2016

[5] Sumit Das, Aritra Dey, Aakash Pal & Nabamita Roy, "Application of Artificial Intelligence in Machine Learning" International Journal of Computer Applications (0975 – 8887) Volume 115 –No. 9, April 2015

[6] Theodouli Terzidou, Christina Miliou and Athanasia Sourvinou, "Agent Supported Serious Game Environment" IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 9, NO. 3, July-September 2016

[7] Xinpin Liu, Liang Zhou, Xingqi Liu, Baozong Yuan Institute of Information Science, School of Computer and Information Technology Beijing Jiaotong University "*Creating Real-time Intelligent Interactive Virtual Environment*" International Conference on Intelligent Computation Technology and Automation, 2015

Technical Discussion Support System

Ankit Agrawal [1], Atharv Dhoble [2], Sudeept Dubey [3], Mrs. Pranjali Kasture [4].
Department of Information Technology, Thakur College of Engineering and Technology, Kandivali (East), Mumbai -400101
aavikas01@gmail.com [1], atharvdhoble@gmail.com [2], sudeeptdubey1997@gmail.com [3], pranjali.kasture@thakureducation.org [4].

Abstract **- In this technological era, the use of advance technologies are increasing day by day. Each and every day new discoveries are taking place which are satisfying the user's need. The main aim of the project is to provide an online platform for various users through which they can discuss and communicate their thoughts and ideas and thus transfer of knowledge would also take place. Technical Discussion Support System encourages reflection and deeper thinking and also focusses on peer to peer communication among the members in the form.**

Keywords—Web Technology, PHP, JavaScript

I. INTRODUCTION

Nowadays e-learning platform is becoming more and more famous. The project is also towards e-learning. It's a PHP based project named "Technical Discussion Support System" is the place where persons after creating their account can share their ideas and make discussions on various topics. In this project CSS, JavaScript, BootStrap and etc are used. This project aims to increase the interest of users regarding any particular topics. This project is on the perspective of students. Expert thinking about the topic can be modelled and learned and useful information is shared among the users. Database connectivity is also used which records the activity performed by the member in the forum. Discussion System motivates learners since active discussion forum have every minute new content and thus can be more would become more attractive to the users.

II. OBJECTIVE

A discussion system contributes to the development of an individual's cognitive and critical thinking skills, allows time for thoughtful, in-depth reflection on course topics, facilitates exploratory learning by allowing you to review and respond to the work of others and paves the way for you to approach your own learning in diverse ways. It provides a platform for users to spread knowledge and also provides an effective communication and discussion among the users. Effective forum topics are open-ended and designed to encourage students to take a position on issues.

III. BACKGROUND

In this technological era, the use of advance technologies are increasing day by day. Each and every day new discoveries are taking place which are satisfying the user's need. The main aim of this project is to provide an online platform for various users through which they can discuss and communicate their thoughts and ideas and thus transfer of knowledge would also take place between them. Online discussion system encourages reflection and deeper thinking and also focuses on peer to peer communication between the members in that forum. The main motive is to provide a wide platform for the users to explore the field of transfer of knowledge through which they would build interest in various topics and can discuss and communicate among themselves related to that particular topic. In future, some features will be added in the discussion forum to overcome the drawbacks in the previous year research papers.

IV. LITERATURE SURVEY

During this phase certain research papers are studied and some of the key findings, research gapes are obtained which can be overcome during the implementation of the project.

Neil Harris and Maria Sandor Developing online discussion forums as student centred peer e-learning environments. It deals with peer based approaches to learning in the e learning environment. Use of discussion board as an assessment. The gap which is observed is that it is limited to use by e learning environment and can be used in other fields also.

Alabo .H. Biriyai and Emmah .V. Thompson Online Discussion Forum a tool for effective teacher student interaction. They used PHP and javascript for creation of the forum. Moderators are used to grant access to the post and are also used to respond to the user's questions and control he content of the discussion forum. It is only limited for teacher and student interaction and is limited to a certain extent.

Ankita Singhal and Brij Mohan Karla Review of online discussion foum in e learning scenario. Web technology is the domain which is used for implementation of the forum. Novek method is proposed to determine post usage to evaluate quality of content in the forum. Removal of error URLs for efficient searching. The paper only suggests the methods of improving forum and does not implement any new feature in the forum.

Jenny McDougall A Study of online discussion forum and needs of adult learners. Concept of authenticity is used. The concept of authentic discussion in classroom study can be replaced by online discussion forum. The paper is only limited to the online discussion at the university level.

Ravi Seethamraju Effectiveness of using Online Discussion Forum for case study analysis. Quantity of responses, Quality of responses and improvement in overall learning. It is reduced structure but a bit complicated to use.

Suryakumari Lane Effective Online discussion forum as an effective learning space. The research methodology is theoretical and empirical. The paper discusses about majority of students having a good learning experience. The author suggest way forward for traditional universities to have blended learning.

V. METHODOLOGY

Technical Discussion Support System focuses on improving cognitive and critical thinking of an individual. Firstly the user needs to login if he/she is the member of the discussion system. If the user is not a member, then he needs to register himself in the forum. Once registered, a wide range of categories is displayed to the user through which he can discuss on his topic of interest. The user can post question and answers and can refer chatbot in case of query. The chatbot consists of predefined set of questions which would help user in case of any issue. User can also read the latest news from the news feed section. Admin has overall control of the activities that is performed by the user. Database connectivity is done which tracks or keeps the record of the activities that are performed by the user in the forum. Web technology is used as a domain and PHP used at front end.

Fig 1shows the flow of the system and various parameters that are considered in the system. Fig 2 displays the use case diagram of the system which involves various actors in the project along with their respective roles. Fig 3 represents the dataflow diagram of the admin which consists of various functions carried out by admin. Fig 4 represents the dataflow diagram of the question module. Fig 5. Represents the login module through which the user can access the system. Fig 6 shows the dashboard which consists of the pending questions. Fig 7 has the record of the user information which are the part of the forum. Fig 8 represents the various categories in the discussion forum. Fig 9 shows the questions that are asked by the registered user in the forum.

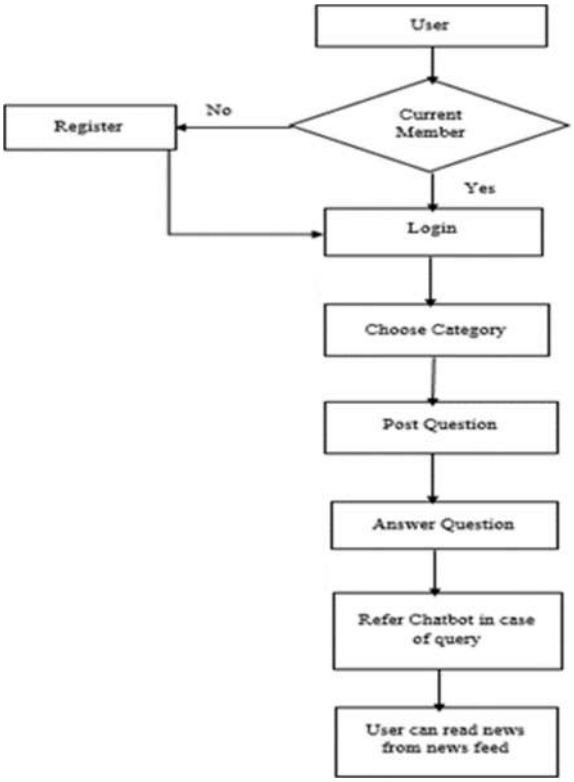


Fig 1: Flowchart of Technical Discussion Support System

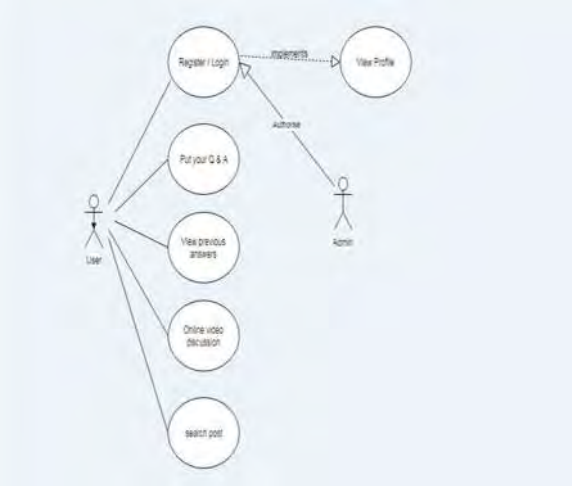


Fig 2: Use case diagram

The Use Case diagram consists of two actors and their respective functionalities.

User- User acts as an actor who firstly register via login and then become a registered user in the discussion forum. Following are the activities that are performed by user as an actor:

1. Register and login
2. Post Questions and answers
3. View Answers
4. Online video discussion
5. Search posts

Admin- Admin takes the control of the discussion forum.

1. Admin can check the question answers posted by user.
2. Has right to accept the post by the users or not

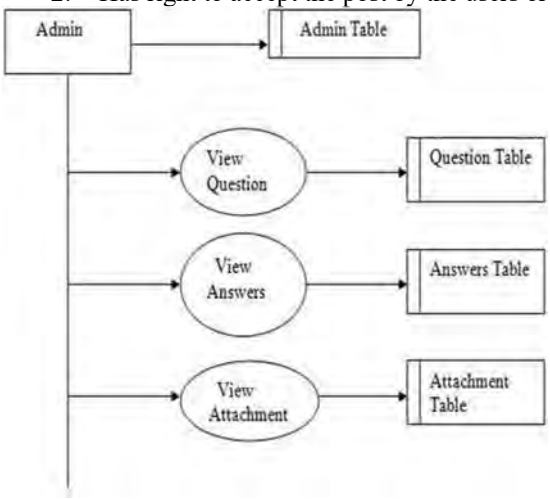


Fig 3: Block DFD diagram of admin module

The admin has the right to view the questions that have been asked, view the answers and also have a role in which he would decide whether the post is relevant or not. The questions that have been asked and answered is stored in the database.

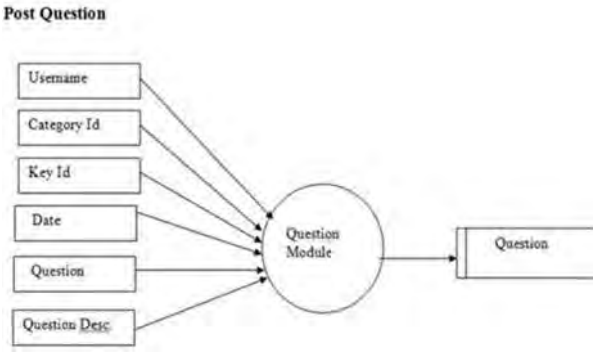


Fig 4: Block diagram of question module

Various category Ids are given based on the user's interest. When the member posts the question then his username ,date on which the question is posted is mentioned in the discussion forum.

VI. RESULT

Technical Discussion Support System focuses on providing an e learning platform for the users for interaction purpose. The main motive behind this project is to establish a knowledge transfer among the people. Firstly, the user needs to register in to the system to be the part of the forum. Once the user is registered a login page is generated which asks user to enter their email id and password for authentication purpose. Some strong crypto algorithms would be applied to ensure security of the technical discussion support system. Once the user would become the member of the forum then he has a right to post questions and answer related to the topic of interest and thus an effective interaction would be established among the members of the forum. News feed section is also provided through which the user would be aware of the daily happenings that are going all around in the world. Chatbot mechanism is going to be incorporated. If a user is having certain doubts while using the technical discussion forum, he can get it cleared via chatbot mechanism which consists of predefined sets of questions along with the answers.

Fig 5: Login Module

The figure represents the login module for the user once he becomes the registered member of the discussion system. The login module consists of email id and password for authentication purpose. Once the user is logged into the forum he can discuss about a particular topic either by posting question or answering the questions.

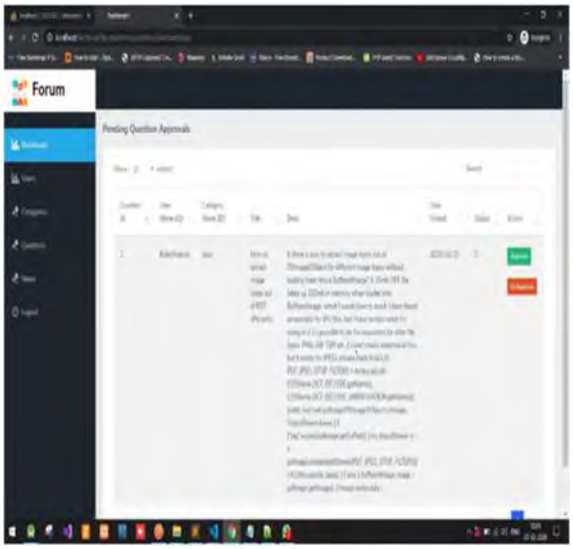


Fig 6: Dashboard Portal

The figure represents the dashboard portal of the discussion system. It consists of some of the pending questions asked by the users along with the user name, category id, title and the date on which it was posted. The admin can take action whether to approve or unapproved.

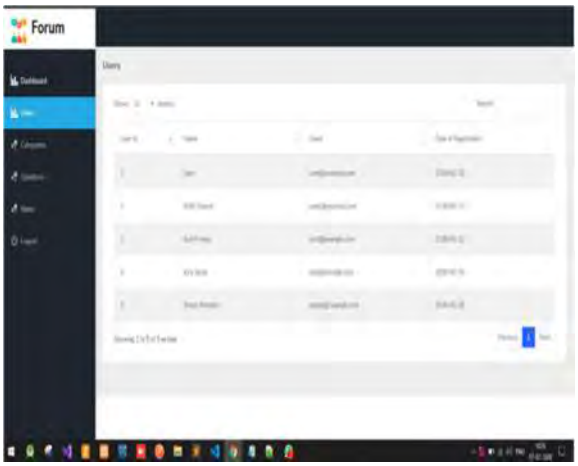


Fig 7: User Information Portal

The figure represents the information of the registered users that is stored in the technical discussion system. The information portal consists of the user id, name of the user, email address of the registered user and the date of registration of the user.

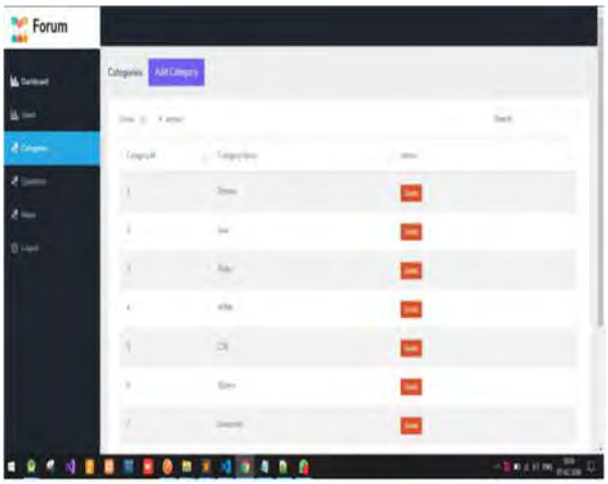


Fig 8: Category Portal

The category portal consists of various categories in the technical discussion system. It has category id, Category name and the action that needs to be taken. Some of the category names are Python, Java, Ruby, HTML, CSS, JQuery, JavaScript.

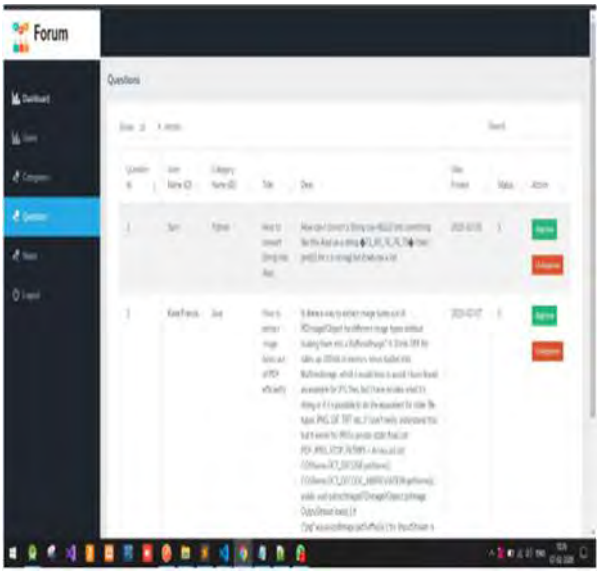


Fig 9: Question Portal

The above figure represents the question portal in the technical discussion support system. The question portal consists of the questions that are asked by the user in the forum. The format of the question portal has question id, user id i.e. the id of user who post question, category id, title, date on which the question is posted along with the status. The admin has right whether to approve the question posted by the user or to unapproved the post.

VII. CONCLUSION

Technical Discussion Support System acts as an interactive tool for knowledge sharing purposes. PHP and database connectivity is used and thus provide a wide platform for the users to explore the field of transfer of knowledge through which they would build interest in various topics and can discuss and communicate among themselves related to that particular topic. News feed section is also provided through which the user would be aware of the daily happenings that are going all around in the world. Chatbot mechanism is going to be incorporated. If a user is having certain doubts while using the technical discussion forum, he can get it cleared via chatbot mechanism which consists of predefined sets of questions along with the answers. A discussion system contributes to the development of an individual's cognitive and critical thinking skills, allows time for thoughtful, in-depth reflection on course topics, facilitates exploratory learning by allowing you to review and respond to the work of others and paves the way for you to approach your own learning in diverse ways. It provides a platform for users to spread knowledge and also provides an effective communication and discussion among the users. Effective forum topics are open-ended and designed to encourage students to take a position on issues.

VIII. FUTURE SCOPE

Further planning involves introduction of video function and Google voice assistance in project so users can discuss the things online and it would be more convenient and working to make an app for it. Our aim is to make the application on current situations and demand.

IX. ACKNOWLEDGEMENT

The quality and ultimate outcome of this paper involved a great deal of support and expertise from many people and we are incredibly privileged to have successfully completed the paper. We take this opportunity to thank everybody for their support.We sincerely thank our principal Dr. B. K. Mishra, vice Principal, Dr. Deven Shah, Mentor Dean Dr. Kamal Shah and HOD Dr. Rajesh Bansode for always guiding us to paths that were right for us. We are grateful to our guide Mrs. Pranjali Kasture who supported and constantly monitored this work and helped us in not only completing this project but also providing us with basic knowledge with their practical aspect that really proved beneficial to us. We are forever thankful and fortunate enough to receive constant encouragement, support and guidance from all IT department faculties that helped us finalize our paper work successfully. We would also like to extend our sincere gratitude to all the laboratory support staff for their timely support. We would like to convey our heartfelt thanks to our parents for their kind cooperation and encouragement without which none of this was possible.

References

[1]. Neil Harris and Maria Sandor; 'Developing online discussion forums as student centred peer e-learning environments', Journal of ascilite Singapore,2015.

[2]. Alabo.H. Biriyai and Emmah .V. Thomas , 'Online Discussion Forum: A Tool for effective student teacher interaction',International Journal of Applied Science-Research and Review,2014.

[3]. Ankita Singhal and Brij Mohan Karla, 'Review of Online Discussion Forum in e-learning scenario', International Journal of Advanced Research In Computer Engineering and Technology,2012.

[4]. Jenny McDougall , 'A study of an online discussion forum and the needs of adult learners' . Australian Journal of Adult Learning Volume 55, ,2015.

[5]. Ravi Seethamraju, 'Effectiveness of Using Online Discussion Forum for Case Study Analysis'.Hindawi Publishing Corporation Educational Research International Volume, 2014.

[6]. Suryakumari Lane, 'Effective Online Discussion Forums as a legal learning space', International Journal of Learning Science, 2016

Interactive paper - Paper display using Projection Mapping

Krishna Madani, Shivani Nawal, Nandan Maurya, Kamal Shah
Thakur College of Engineering and Technology, Mumbai, India
krishnamadani@gmail.com, shivani.nawal1@gmail.com, nandanmaurya209@gmail.com

Abstract— **With advancement in technology especially in the field of virtual reality the user experience has become easy, interactive and interesting. We here are going to present the concept of Interactive paper i.e. Paper display using projection mapping. The traditional way calibrates the projector to display the content. Calibration is the process to setting up the connection between the device that projects and the one on which it is projected. The overall process of calibration consumes a lot of time and if there is a little mishandling it can cause disturbance in the projection. Our project overcomes this issue. Our project's main aim is to provide an interactive platform for users to implement and understand things in a better way. The Object that is to be interacted with is 3D designed in a Virtual environment with the help of computer software like Unity, AutoCAD etc. It is then projected on the paper and mapped accordingly by means of specialized software. This is where the Projection mapping is implemented as well as restricted because of required stillness of real object but it is extended further printing a pattern on the paper and tracking it by software by means of a camera. So real-time scenes of the paper are fed to the software by the camera and accordingly, updates are made if the paper is moved from its original position. This concept has been in existence from past 3-4 years, but it is not implemented or used widely. This concept models the e -light emitted by projectors and reflected into cameras, and to consider the displayed content as additional information useful for direct alignment and to provide the projection accurately.**

The projection mapping on paper is going to be achieved by combining hardware with software to produce the result accuracy. We here are displaying the pattern on the paper in 3D format to make it interactive. This can be widely used in the field of education and medical science.

Key words: Projection Mapping, Augmented Reality, VR.

I. INTRODUCTION

Nowadays Virtual reality and Augmented reality have taken over the world. With these two powerful field of domains we are able to experience a lot of new things. The most common examples of these are : broadcasting messages through VR , Playing sports game is the most trending one , Pokémon Go game is one the examples for AR. The point here is that these fields are applied to the gaming and entertainment purpose so effectively why not use it for Educational purposes.

In this project projection of objects / patterns are done on the paper. The projection of Virtual objects will be done on a sheet of paper which will be having a specific pattern printed in it for tracking i.e. by tracking the pattern we can have the virtual effect of the object/pattern. Increase in technology is rising highly and with them new concepts need to be adapted and implemented to efficient use. This concept is an upcoming technology in the field of virtual reality which can be helpful in various fields. In the end this project is just a start towards the new technology to be used in education institutes and medical science.

In education there are various ways such as explaining science topics to students so that their understanding of subjects improves. In medical science Doctors can explain to students as well as the patients each part of body in 3D, which can actually give the feeling of an actual object.

II. OBJECTIVES OF THE PROJECT

The aim is to make the conventional paper using a unique experience by interacting with it with the help Augmented Reality and a concept known as Projection Mapping.

1. The Object that is to be interacted with is 3D designed in a Virtual environment with the help of computer software like Unity, AutoCAD etc.
2. It is then projected on the paper and mapped accordingly by means of specialized software. This is where the Projection mapping is implemented as well as restricted because of required stillness of real object but it is extended further printing a pattern on the paper and tracking it by software by means of a camera.

So real-time scenes of the paper are fed to the software by the camera and accordingly, updates are made if the paper is moved from its original position

III. RELATED WORK

There has been vast work on permitting augmented truth eventualities using projection mapping. Whilst recent work has captured lots press interest, there has been much work over the last many years. Since this early paintings, which explored the use of big and

curved projection surfaces that would be semi-automatically calibrated, there was tons work on projection systems, calibration strategies, and camera-projector structures. Systems have looked at the problem of projecting and interacting across more than one flat surfaces, even the usage of a moving hand-held projector. One particular trouble explored is the mixing of a couple of projections on flat or pseudo-flat surfaces. These systems borrow from the big literature on image-based totally rendering and mixing techniques. However, these structures have no longer explored complicated geometry or moving surfaces. More recently, with the arrival of actual-time depth cameras, projector systems have all started to cope with greater complex geometries. IllumiRoom explored projecting across the periphery of a tv display, and RoomAlive explores multiple projections inside an entire room. However, neither of those systems deal with moving projector/cameras or moving scenes. One instance of a transferring hand-held projector gadget which is used to render content onto real-world geometry of arbitrary form, captured the usage of the KinectFusion gadget. However, this machine did no longer scale to transferring gadgets, nor more than one projectors. Other work has focused at the problem of correcting the projected photograph because of inherent shade and visual artifacts the usage of various radiometric repayment strategies, or by using modeling the inverse light shipping of the projector. Some projector structures compensate in actual-time for shade modifications on dynamic projection surfaces. Other structures compensate for environment lighting fixtures modifications or maybe the cloth houses of the scene. These structures but frequently do now not deal with complicated geometries or assist motion of the floor or projector/digicam device. However, a time-eating static calibration approach changed into finished by manually moving a go hair within the projector view to highlight pixels that light up known object factors. In comply with-up work, used fiber optic sensors embedded inside the object to localize a shifting object, but now not in actual-time. There has been comparable research which demonstrated projecting on a moving item, however required infrared markers and/or magnetic sensors at the object, and best a small working extent. DisplayObjects extensively utilized a marker-based totally monitoring gadget, to music the display objects. In those structures the projector and sensor device are assumed to remain static. Researchers supports a moving projector digital camera system, however does not scale to a couple of projectors modeling the light shipping to make sure accurate in keeping with-pixel projection rendering. Researchers have offered a projection mapping gadget this is able to dynamically task onto moving objects. They use a 1000fps digital camera and galvanometer mirrors to tune and illuminate the object with a single projector. The ensuing monitoring is extraordinarily fast and dazzling effects are performed. Note that this monitoring method is orthogonal to our contribution and could be used to seriously reduce the latency of our machine.

IV. BACKGROUND

PROJECTION MAPPING

Projection mapping, similar to video mapping and spatial augmented reality, is a projection technique used to turn objects, often irregularly shaped, into a display surface for video projection. These objects may be complex industrial landscapes, such as buildings, small indoor objects or theatrical stages. By using specialized software, a two- or three-dimensional object is spatially mapped on the virtual program which mimics the real environment it is to be projected on. The software can interact with a projector to fit any desired image onto the surface of that object. This technique is used by artists and advertisers alike who can add extra dimensions, optical illusions, and notions of movement onto previously static objects. The video is commonly combined with, or triggered by, audio to create an audio-visual narrative.

AUGMENTED REALITY

Augmented reality (AR) is an interactive experience of a real-world environment where the objects that reside in the real world are enhanced by computer-generated perceptual information, sometimes across multiple sensory modalities, including visual, auditory, haptic, somatosensory and olfactory. AR can be defined as a system that fulfills three basic features: a combination of real and virtual worlds, real-time interaction, and accurate 3D registration of virtual and real objects. The overlaid sensory information can be constructive (i.e. additive to the natural environment), or destructive (i.e. masking of the natural environment). This experience is seamlessly interwoven with the physical world such that it is perceived as an immersive aspect of the real environment. In this way, augmented reality alters one's ongoing perception of a real-world environment, whereas virtual reality completely replaces the user's real-world environment with a simulated one. Augmented reality is related to two largely synonymous terms: mixed reality and computer-mediated reality.

The traditional and mostly used method to experience the VR is via the Head Mounted Devices and if we are talking sharing this experience with Group of people than it would be CAVE or using the PC so as to view the orthographic projection of VR. In both the cases user has to wear some type of gadget / device to operate and sharing it would mean that many devices and complexity or only one person controlling the environment and other are just passive viewers which is not very interactive and feature rich method. So as to provide the user/s a featured experience without any need of wearable devices and to be able to share this with a group with ease, this system is proposed.

V. PROPOSED SYSTEM

Purpose –The aim is to make the conventional paper by interacting with it with the help Augmented Reality and a concept known as Projection Mapping.
Process –Paper is tracked, via the pattern made on it, by a camera and feedback is provided to the application which calculates the relative coordinated with respect to the surface on which paper is to be placed and mapped to the correct content and projected so as to give a feeling of a interactive paper display. So real-time scenes of the paper are fed to the software by the camera and accordingly, updates are made if the paper is moved from its original position.

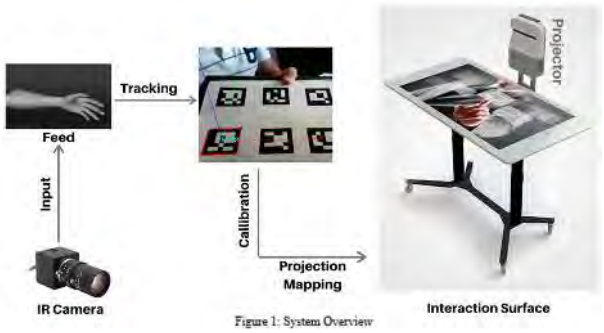


Figure 1: System Overview

VI. METHODOLOGY

Our project is AR based and involves both hardware and software. This project combines both the technologies of projection mapping and augmented reality to get the results. The following are the steps involved to get the result.

Step 1: The paper is kept on the surface which on which the projected content will be displayed.

Step2: Using the marker technology and camera the coordinates of paper are captured.

Step 3: This is then sent to the system to process the data and then it is projected.

Step 4: When the user moves the paper the camera again captures the coordinates again as a feedback and then accordingly it changes the projection.

VII. FLOWCHART

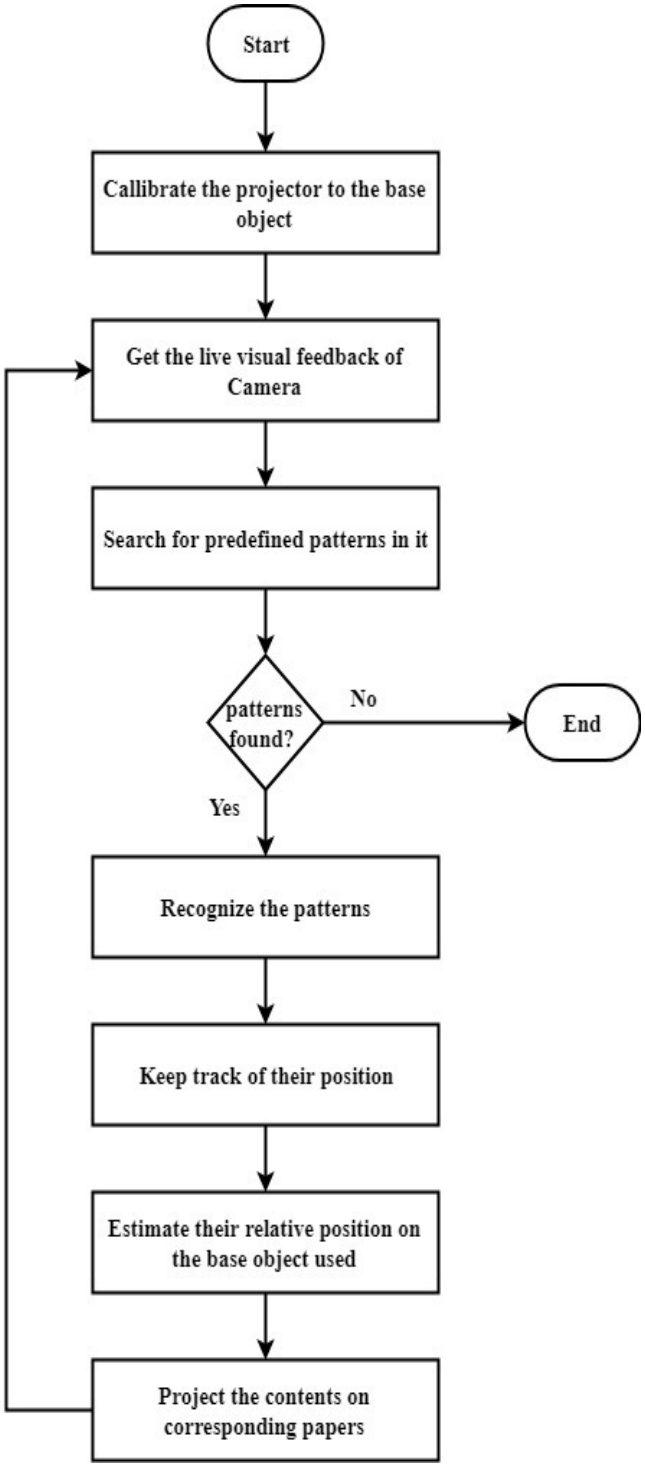


Figure 2: Flow of the System

VIII. RESULT

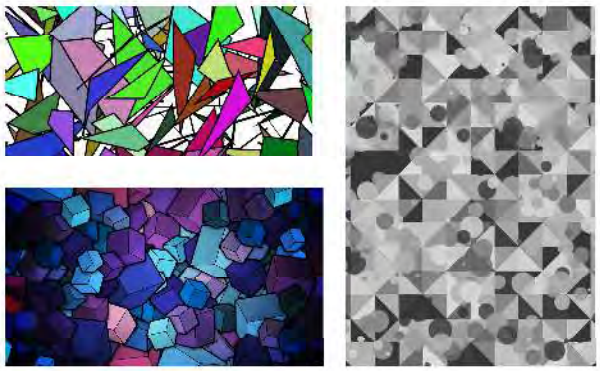


Figure 3: Markers to be tracked

The above image has the markers which are going to be tracked with the help of a software. These patterns are numbered and their relative distance is calculated.

Figure 4: tracking the markers

The application on start searches for the marker printed on the paper. It detects it gives them numbering. After the numbering is done it tracks the marker when it is moved.

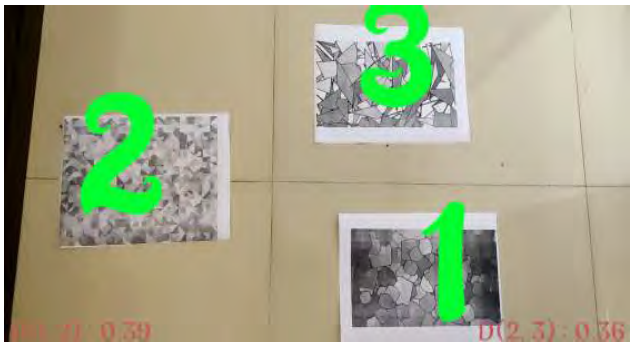


Figure 5: Showing distance between papers

Figure shows a sample testing program which tracks several markers printed on papers and calculated a relative difference of positions in terms of actual distance which will be used for calibrating and mapping the projection accordingly. Here, the distance between marker 1 and 2 is 0.39 and that of 1 and 3 is 0.36

IX. CONCLUSION

Our project is an AR system that uses the concept of projection mapping to display the object on a paper. The Object that is to be interacted with is 3D designed in a Virtual environment with the help of computer software like Unity, AutoCAD etc. It is then projected on the paper and mapped accordingly by means of specialized software. This is where the Projection mapping is implemented as well as restricted because of required stillness of real object but it is extended further printing a pattern on the paper and tracking it by software by means of a camera. Real-time scenes of the paper are fed to the software by the camera and accordingly, updates are made if the paper is moved from its original position. Here we will be able to make a fully efficient system when the implementation starts as of now we have just completed 2 phases of the waterfall model i.e. requirements gathering and system design.

X. FUTURE SCOPE

This project has tremendous development scope which is only limited by a person's imagination.

It can be improved and used in various ways such as:-

1. The pattern on the paper can be made with infrared ink so that the paper can be tracked in low luminous conditions such as closed areas or in the night and it also improves the contrast of the projection.
2. Hand or Body tracking machines can be used to implement body gestures, so as to better interact with the projection and make the experience as equipment free and natural as possible.

XI. ACKNOWLEDGMENT

We sincerely thank our guide Dr. Kamal Shah for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Rajesh Bansode, the Principal, Dr. B. K. Mishra, and the college management for their support.

References

Web References –

[1]http://projectinstinct.capsciences.net/?q=node/34

[2] https://youtu.be/bEDDj7Ahgwo

[3]https://www.youtube.com/watch?v=bEDDj7Ahgwo

[4]https://team.inria.fr/potioc/scientificsubjects_old/par/

Journal References –

[1] Martin Hachet , Jeremy Laviole ,Jean Baptisted and Aurelie Cohe , "Touch based interfaces for interacting with 3D content in public exhibitions" , IEEE Computer Graphics and Applications , March 2013

[2] Kerstin S. Haring , Nathan L. Tenhundfeld , Victor Finomore , Mormon Redd and Brian Tidball , "Analysis of Using Virtual Reality (VR) for Command and Control Applications of Multi-Robot Systems", Human-Robot Interaction Companion, March 5–8, 2018.

[3] AN Heng, LI Fangfang, YAN Guangrong and LEI Yi , "Research on 3D Dimension Automatic Projection and Quick Update", IEEE Computer Graphics and Applications , 2009

[4] Samuel Audet and Masatoshi Okutomi , " A User-Friendly Method to Geometrically Calibrate Projector-Camera Systems" , Tokyo Institute of Technology 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan,2009

[5] Tyler Jhonson and Henry Fuchs ,"Real-Time Projector Tracking on Complex Geometry Using IEEE Computer Society Conference on Computer Vision and Pattern Recognition · June 2007

[6] T. Senoo, Y. Yamakawa, Y.Watanabe, H. Oku, andM. Ishikawa, "High-speed vision and its application systems," *Journal of Robotics and Mechatronics*, vol. 26, no. 3, pp. 287–301, 2014.

[7] D. Kobayashi and N. Hashimoto, "Spatial augmented reality by using depth-based object tracking," in *Proceedings of the ACM Special Interest Group on Computer Graphics and Interactive Techniques Conference (SIGGRAPH '14)*, Vancouver, Canada, August 2014.

[8] R. Koizumi, D. Kobayashi, and N. Hashimoto, "Acceleration of dynamic spatial augmented reality systemwith a depth camera," in *Proceedings of the International Conference on Cyberworlds (CW '15)*, pp. 50–53, October 2015.

[9] Y. Zhou, S. Xiao, N. Tang, Z. Wei, and X. Chen, "Pmomo: projection mapping on movable 3D object," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 781–790, Santa Clara, Calif,USA,May 2016.

[10] J. Knibbe, H. Benko, and A. D. Wilson, "Juggling the effects of latency: software approaches to minimizing latency in dynamic projector-camera systems," in *Proceedings of the 28th Annual ACMSymposium onUser Interface Software&Technology (UIST'15)*, pp. 93–94, ACM, Daegu, South Korea, November 2015.

Experimental Analysis Of Agricultural Data Using Data Mining

Manvi Jain [1], Jitendra Patel [2], Mansi Pawar [3].
Information Technology
Thakur College of Engineering and Technology
Mumbai, India
manvijain231998@gmail.com [1], jitu166.jp@gmail.com [2], mansipawar51@gmail.com [3]

Abstract— **Data Mining is a technique which focuses on large data sets to extract information for prediction and discovery of hidden patterns. Data Mining is applicable on various areas like healthcare, insurance, marketing, retail, communication, agriculture. Agriculture is the backbone of country's economy. It is the important source of livelihood. Agriculture mainly depends on climate, topography, soil, biology agricultural. The current study presents the different data mining techniques and their role in context of soil fertility, nutrient analysis. Decision tree is a well-known approach for classification in data mining. C4.5 and Classification and Regression Trees (ID3) are two widely used decision tree algorithms for classification. The main drawback of C4.5 algorithm is that, it is biased towards attributes with more values while ID3 algorithm produces misclassification errors when the domain of the target attribute is very large. In view of these limitations, this paper presents a modified decision tree algorithm. The model is tested with test data set of soil samples. The test proves that the modified decision tree algorithm has higher classification accuracy when compared to C4.5 and ID3 algorithms. Classification of soil is the separation of soil into classes or groups each having similar characteristics and potentially similar behavior. Classification of soil is needed so that farmer can know the type of soil and can plough the crops depending on the type of soil.**

Keywords— DecisionTree, C4.5,ID3

I. INTRODUCTION

Globally, day to day the need for food is escalating, therefore the agricultural scientists, farmers, government, and researchers are trying and using varied techniques in agriculture for improvement in production. As an impact, the information generated from the agricultural data is increasing day by day. Because the volume of information enlarged, it needs a spontaneous method for this information to be mined and analyzed once required. Data Mining techniques will be used for prediction the longer term trends of agricultural processes.

Data Mining techniques are of two types, one is descriptive which considers the existing data and another is predictive which depends on probability for future analysis. Data Mining process involves

a) Collect, clean and load the data into data warehouse system
b) Stores the data in multidimensional format
c) Provides information access to analysts and decision makers
d) Analyzation of data using different applications
e) Presents the data using different patterns

II. BACKGROUND

Agricultural researchers and farmers deploy sensors at their remote agricultural-fields to obtain the data of temperature, humidity, soil moisture and so on. Automatic collection of those data greatly helps their analytical works. Currently, they are relying on network providers, for example, cellular phone network to achieve such automatic collection from their remote sites, which is not feasible for most of the farmers due to operational cost. The review and meta-analysis of yield data comparing organic and conventional agriculture showed that currently organic yields of individual crops are on average 80% of conventional yields The analysis of 362 datasets also showed a high variation of the yield gap of organic agriculture (standard deviation 21%). Some of this variation seems systematic.E.g. soybean, some other pulses, rice and corn score higher than 80% and wheat, barley and potato scoring lower than 80%. Most regions have relative yields fairly close to the overall average.

III. PROBLEM DEFINITION

In current scenario, there are no such recommendation system available online. And also farmer has to physically go and get recommendation based on their soil report by quality executives. In this project, we've got suggested an analysis of the soil information using Decision Tree algorithms and prediction technique. By using Decision tree algorithm, we recommend farmers about their field soil quality and suggest crops which are suitable to grow in that soil.Use of information technology in agriculture can change the scenario of decision making and framers can yield in a better way.

IV. OBJECTIVE

The objectives of this study were to characterize and classify the soil to provide more details about the quality information of soil. An additional objective of this study was to seek out approximate crops can be growing on that field. This study were to characterize and classify the soil to provide more details about the subsurface morphological information. An additional objective of this study was to seek out an approximate value of soil loss from the Lower Moshi Irrigation Scheme using the Universal Soil Loss .

V. LITERATURE SURVEY

We have undergone certain research papers based on the Experimental Analysis Of agricultural Data Using DataMining and found out certain methodology and key findings. During the process, we have .also identified certain research gaps which we can overcome while implementing our project.

Ref no.	Paper Title	Author	Year Of Publication	Key Findings	Research Gaps
1	Analysis Of agriculture data Using Data Mining	Lucinda Edwards & Carolina Eriksson	2017	Various data mining techniques are implemented on the input data to assess the best performance yielding method. The present work used data mining techniques PAM, CLARA and DBSCAN to obtain the optimal climate requirement of wheat like optimal range of best temperature, worst temperature and rain fall to achieve higher production of wheat crop.	Clustering is considered as an unsupervised classification process. A large number of clustering algorithms have been developed for different purposes. Clustering techniques can be categorised into Partitioning clustering, Hierarchical clustering, Density-based methods, Grid-based methods and Model based clustering methods.
2	Analysis Of Soil Behaviour and Prediction of Crop Yield using data mining approach	Supriya Dm , Karnataka	2016	The researcher express that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantages.	The researchers implemented K-Means algorithm to forecast the pollution in the atmosphere, the K Nearest Neighbour is applied for simulating daily precipitations and other weather variables and different possible changes of the weather scenarios are analyzed using Support Vector Machines.
3	A Survey On Predictive Analysis in	A. Agrawal J. Basak	2018	Soil classification was measured serious to study due to depending upon	suggested an analysis of the soil data using different
	Agricultural soil data to predict the best fitting crop	V. Jain R. Kothari		the fertility class of the soil domain knowledge experts determines which crops should be taken on that particular soil and which fertilizers should be used for the same.	algorithms and prediction technique.
4	Data Mining Technique to Predict the Accuracy Of the Soil fertility	Dr. S. Hari Dr.Jayashree	2017	The research studies on application of data mining techniques in the field of agriculture. Some of the techniques, such as ID3 algorithms, the k-means, the k nearest neighbour, artificial neural networks and support vector machines applied in the field of agriculture were presented.	This article explores the applications of data mining techniques in the field of agriculture and allied sciences. Historical crop yield information is important for supply chain operation of companies engaged in industries that use agricultural produce as raw material.
5	Data Mining in Agriculture prediction soil fertility	Dr.Shrinivashan , Dr.Mailthili M.	2017	A soil test is the analysis of a soil sample to determine nutrient content, composition and other characteristics. Tests are usually performed to measure fertility and indicate deficiencies that need to be remedied.	In this paper demonstrated acomparative study of varied classification algorithms i.e. Naïve bayes, J48 (C4.5), JRip with the assistance ofdata mining tool .J48 is incredibly easy classifier to form a decision tree.

Fig: Literature Review Table

VI. METHODOLOGY

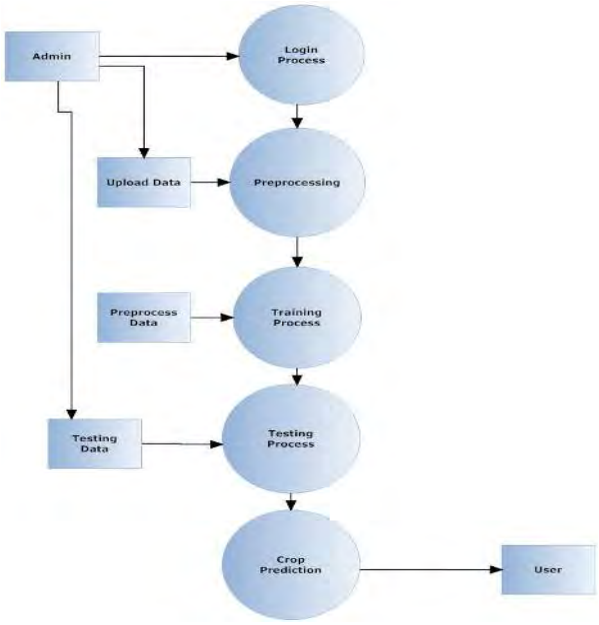


Fig. Methodology Block Diagram for Analysis of Agricultural Data using data mining

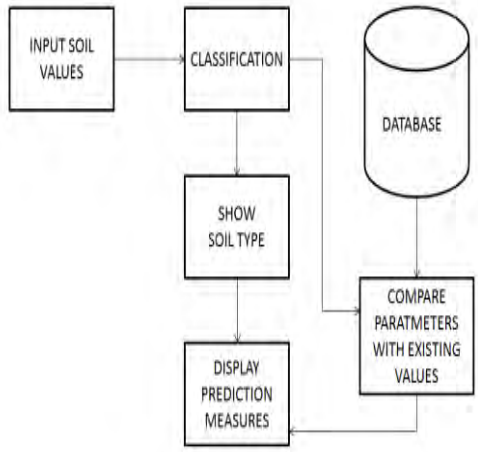


Fig: Soil Classification

VII. ALGORITHM IMPLEMENTED

Decision Tree Algorithm

In this project, we've got suggested an analysis of the soil information using Decision Tree algorithms and prediction technique. By using Decision tree algorithm, we recommend farmers about their field soil quality and suggest crops which are suitable to grow in that soil.

Regression Tree Algorithm

It models a target prediction value based on independent variables. ... Linear **regression** performs the task to predict a dependent variable value (y) based on a given independent variable (x).We have used the

algorithm to predict rainfall based on the previous year rainfall data.

The main drawback of C4.5 algorithm is that, it is biased towards attributes with more values while ID3 algorithm produces misclassification errors when the domain of the target attribute is very large**.**

VIII. RESULT AND DISCUSSION

This project is fundamentally related to use of information technology in agriculture can change the scenario of decision making and framers can yield in a better way. For decision making on overall issues related to agriculture field; data mining plays a vital role. The survey discussed the role of data mining in terms of agriculture field. We have also discussedvarious types of soils, several data mining techniques in agriculture and soil containment.

IX. CONCLUSION

In this project, we've got suggested an analysis of the soil information using Decision Tree algorithms and prediction technique. By using Decision tree algorithm, we recommend farmers about their field soil quality and suggest crops which are suitable to grow in that soil. Agriculture is the utmost important area especially in the mellowing country like India. Use of information technology in agricultre can change the scenario of decision making and framers can yield in a better way. For decision making on overall issues related to agriculture field; data mining plays a vital role. The survey discussed the role of data mining in terms of agriculture field. We have also discussed various types of soils, several data mining techniques in agriculture and soil containment.

X. FUTURE SCOPE

The system “crop prediction using data mining technology” is developed and tested successfully and satisfies all the requirement of the client. The goals that have been achieved by the developed system are: • Simplified and reduced the manual work. • Large volumes of data can be stored. • It provides Smooth workflow. We can add device to get values directly from soil testing lab to server. We can add module if any queries is there, the staff can directly interact with the administrator very easily.

XI. ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have completed the project successfully. I would like to thank everyone for their guidance.

We sincerely thank our principal, Dr.B.K.Mishra , Vice Principal Dr. Deven Shah and HOD Dr.Rajesh Bansode for always encouraging us to do our best. We highly indebted to our guide **Mr.Vijay Kumar Yele,** who supported and constantly supervised us throughout the project and helped us in not only completing this project but also provided us with ample amount of knowledge that was really beneficial to us.

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of IT Department who helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in laboratory for their timely support.

We would like to express our gratitude towards our parents for their kind co- operation and encouragement which helped us in completion of this project.

References

[1]. Ahmed, A.U., R.V. Hill, L.C. Smith, and D.M. Wiesmann. 2007. The World’s Most Deprived: Characteristics and Causes of Extreme Poverty and Hunger.

[2]. Allen, J.C., and N. Mar. 1985. Soil Response to Forest Clearing in the United States and the Tropics : Geological and Biological Factors Soil Response to Forest Clearing in the United States and the Tropics : Geological and Biological Factors1. 17(1): 15–27.

[3]. Andersson, J. a., and S. D’Souza. 2014. From adoption claims to understanding farmers and contexts: A literature review of Conservation Agriculture (CA) adoption among smallholder farmers in southern Africa. Agric. Ecosyst. Environ. 187: 116–132Available at http://linkinghub.elsevier.com/retrieve/pii/S0167880913002685 (verified 28 May 2014).

[4]. Anthofer, J. 2004. The Potential of the SRI for Poverty Reduction in Cambodia. Consultancy Report. FSNPSP

[5]. Chapagain, a. K., and a. Y. Hoekstra. 2010. The green, blue and grey water footprint of rice from both a production and consumption perspective. Ecological Economics(40): 1–62.

Crime Prediction And Analysis

Dona Darji, Himani Mehta, Priyanka Iyer, Hetal Amrutia
Information Technology Thakur College of Engineering and Technology Mumbai,India
donadarji24@gmail.com, himanimehta20@gmail.com, priyankaiyer98@gmail.com

Abstract—**Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions. In this project, we will be using the technique of machine learning and data science for crime prediction of India crime data set. The crime data is extracted from the official portal of the Indian government. It consists of crime information like location description, type of crime, date, time, latitude, longitude. Before training of the model data preprocessing will be done following this feature selection and scaling will be done so that accuracy obtained will be high. The Naive Bayes classification and various other algorithms will be tested for crime prediction and one with better accuracy will be used for training. Visualization of dataset will be done in a graphical manner and in a way that can be easy to understand for the concerned authorities. This information can be used to determine a number of characteristics such as the time at which the criminal rates are high or at which month the criminal activities are high according to the locality specified. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Instead of focusing on causes of crime occurrence like criminal background of offender, political enmity etc we are focusing mainly on crime factors of each day. The soul purpose of this project is to give a basic idea of how machine learning can be utilised by the law enforcement agencies and intelligent forces to identify patterns in commited crimes, predict crime prone regions and unsafe time periods and hence solve crimes at a much faster rate and thus reduce the crime rate. It is not restricted to our Country; this can be used in other states or countries as well depending upon the availability of the dataset.**

Keywords—Crime analysis, Criminal prediction, Naı̈ve Bayes

I. INTRODUCTION

Violations are the basic peril to humankind. Crimes can't be anticipated since it is neither precise nor arbitrary. Infringement happens from town to enormous urban zones. Infringement are of different sort robbery, murder, ambush, assault, battery, counterfeit confinement, stealing, crime. Since infringement is extending there is a need to light up the cases in a significantly snappier manner. Every day the crime rate is expanding fundamentally. It is growing and spreading at a snappy and colossal rate.

It is the commitment of the police office to control and lessen the crime rates. The desire to commit an offence and criminal unmistakable confirmation are the significant issues to the police office as there are enormous proportion of crime data that exist. There is a need for advancement through which the case settling could be speedier. Conjointly the advances and modern ways encourage culprits in accomplishing their wrongdoings. As indicated by Crime Records Bureau wrongdoings like lawful offense, incendiarism and so on are wilted though violations like homicide, sex misuse, assault and so on are amplified.

Regardless of whether we are not ready to realize who are the exploited people we can in any case anticipate the spot that crime may occur.The anticipated results can not be guaranteed of 100% exactness anyway the outcomes shows that our application causes in lessening rate to a specific degree by giving security in wrongdoing delicate regions.

There should be an ideal crime analysis tool that should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action. However, currently, there are a number of major challenges that are encountered. Increase in the size of crime information that has to be stored and analyzed, Problem of identifying techniques that can accurately and efficiently analyze this growing volumes of crime data, Different methods and structures used for recording crime data, The data available is inconsistent and are incomplete thus making the task of formal analysis a far more difficult. Investigation of the crime takes longer due to the complexity of such issues.

Therefore, there is a need for building a system that can if not overcome, limit the challenges that are faced by the existing systems. The main features it should contain should be identifying crime patterns and analysing them, determining crime prone regions of a specific location, providing information to formulate strategies for crime prevention and reduction. So, for building such an amazing crime examination instrument we've to assemble crime records and judge it. Finding the examples and patterns in wrongdoing is a troublesome issue. To detect an example, crime experts take a lot of your time, looking over data to see whether a particular crime fits into a superior known example. In the event that it doesn't work into a partner degree existing example, at that point the data ought to be named another example. When an analyst works as an example, it will be utilized to foresee, envision and stop crime.

The purpose of this is to make crime gauge using the features present in the dataset. The dataset is

expelled from the official regions. With the help of AI computation, using python as focus we can envision the kind of crime which will occur in a particular territory. The target is to prepare a model for desires. Building the model will be finished using better count depending on the precision. The Naive Bayes course of action and other computation will be used for wrongdoing desires. Impression of dataset is done to separate the infringement which may have occurred in the country. This work helps the law prerequisite workplaces to predict and recognize infringement in the district with improved exactness and thus diminishes the crime rate.

II. BACKGROUND

The job of PCs has been expanded in varying backgrounds from the fund division to general stores. As of late police powers have been improving their customary technique for wrongdoing revealing with new innovative progressions to expand their yield by productively recording violations to help their examination. Information isn't only a record of wrongdoings, it additionally contains significant data that could be utilized to interface wrongdoing scenes dependent on the business as usual of the offender(s), recommend which guilty parties might be liable for the wrongdoing and furthermore distinguish those guilty parties who work in groups (wrongdoer systems) and so forth. In this day and age, computers are assuming a significant job in the examination of a wide range of crime from those that are considered as volume wrongdoing (thievery, vehicle related crimes and so on.) to significant wrongdoing, for example, extortion, tranquilize dealing, murder etc.It isn't a simple errand for a Police expert to physically unwind the inalienable complexities inside police information and this issue is exacerbated when the investigation is attempted by a group. The appropriation of the information to the group may cause huge data, which could be valuable to tackle the violations, to be missed as every part isn't in control of every single pertinent truth. For quite a while, criminologists and analysts have been applying their aptitudes and information attempting to anticipate when and where the following arrangement of wrongdoings will happen, with changing degrees of achievement. The crime rate and the more prominent familiarity with present day crooks put a strain on the current techniques. Human thinking bombs when given a great many records. In this manner, there is obviously a necessity for a toolbox to help with breaking down the information which will utilize constrained assets. Information Discovery in Databases (KDD) procedures can be utilized to uncover information which is past instinct.

III. LITERATURE REVIEW

Crimes are a significant threat to humankind. There are many crimes that happen regularly. Perhaps it is increasing and spreading at a fast and vast rate. Crimes happen from small villages, town to big cities. Crimes are of different type robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasingly there is a need to solve the cases in a much faster way. The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there is a tremendous amount of crime data that exist. There is a need for technology through which case solving could be faster. The aim is to make crime prediction using the features present in the dataset. The dataset is extracted from the official sites. With the help of machine learning algorithms, using python as core we can predict the type of crime which will occur in a particular area. Developing a crime analysis tool using different data mining techniques can help law enforcement department to efficiently handle crime investigation. The proposed tool enables agencies to easily and economically clean, characterize and analyze crime data to identify actionable patterns and trends. The criminal forecast turns out to be significantly more confused when the criminal-related information, for example, criminal profile and criminal interpersonal organization, and the geographic information than can be any information related with the date and area where the incident occurrence, are considered.
episodes, for example, geographic profile, are considered.
[1]. Various criminal activities can show the criminal's characteristics and his/her intentions to a great extent. Hence, analysing the crime and predicting it based on the characteristics of the criminal activity is of paramount importance. Therefore, analysing the criminal's behaviour has been proven to play a significant role in criminal behavior analysis [6]. This research work suggests that linear regression algorithm is the most effective Machine learning algorithm to analyze crime patterns and can work with randomness to an extent as opposed to decision stump algorithms which have a relatively poor performance which only discusses algorithms to analyse crime patterns [5].

IV. OBJECTIVE

As we realize that the quantity of bodies of evidence recorded against various violations have expanded in the previous 3-5 years.There is a need to examine the quantity of wrongdoings carried out, the sort of wrongdoing perpetrated, recognize the locales where there is the most extreme crime rates. There is a need to build up a framework that can anticipate districts which have high likelihood for criminal events and can imagine crime inclined zones. With the expanding appearance of automated frameworks, crime information examiners can help the Law requirement officials to accelerate the way toward understanding violations and can likewise encourage overwhelming watching in those districts as to anticipate wrongdoings.

We can apply distinctive measurable examination calculations to the info datasets i.e, the quantity of crime rates in a specific interim of time, the districts where most extreme crime rates have been recorded and different variables can be considered to foresee

future crimes.Therefore, important advances can be taken to forestall the crimes specifically region.Thus, the goal is to train a model for forecast. The preparation would be finished,utilizing the preparation informational index which will be approved utilizing the test dataset. Building the model will be done

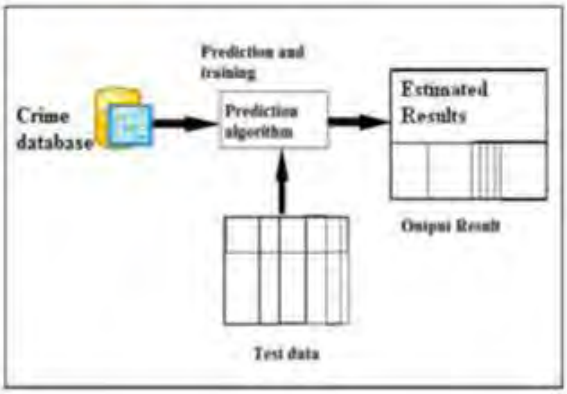


Fig.1 Proposed Architecture

utilizing better calculation relying on the exactness. Representation of dataset is done to break down the wrongdoings which may have happened in the nation.Generated model can be used in many ways and can be integrated with multiple applications like in intelligent investigative systems, crime analysis systems etc or can be used as an independent system. It has a wide range of applications. Resources and funding provided by respective stakeholders can be used in scaling the computing power i.e. renting or purchasing the GPUs, which will help us to develop a better model.

V. METHODOLOGY

Utilizing Machine Learning Technology, different calculations can be utilized to :- 1. Examine criminal patterns, Predict future spots for criminal activity. 2. Determine the perilous time frames of different regions.3. Predict the sort of violations that could occur in various locales dependent on notable information and patterns.4. Come up with measures to counteract those violations.

Python library Scikit-learn (sklearn) is used for preprocessing the dataset. Some attributes in the csv files contain string values and others are numeric values. In order to use this dataset in machine learning models, the text features need to be converted into a numeric value. Python library numpy is used to contain both features and label of the dataset after converting them into numeric values. Attributes with string data type are “Day”, “Category”, “Address” columns. Scikit-learn has a preprocessing package that converts string data into numeric data. This package gives an integer value to each unique item after sorting items in ascending alphabetical order. Datetime attribute is also a string data type, however this is converted into a datetime object and four different attributes are obtained from it: “Hour”,”Date”,”Month” and “Year”.

To avoid overfitting and getting more realistic accuracy, the dataset is divided into two portions: testing dataset and training dataset. Training dataset contains all features along with the target label. Testing dataset only contains the features from which a machine learning model predicts the target label. Scikit-learn’s model_selection package contains a class test_train_split that splits the original dataset into testing and training dataset. The default value of the test dataset size is 25% of the original dataset. This default value is used in the conducted experiments.

While the given features give sufficient information about a crime incident, new features can be extracted from the given features which might prove to be useful.

One assumption is made before using any predictive model is, some features might be more useful than others. While more features describe the data better, too many features can make classification complicated and cause overfitting. sklearn.feature_selection module uses univariate statistical tests to find features that are best related to the target label. select_percentile class of this module takes a percentage input and returns that percentage of best features. For classification problems f_classif function of this class is used. Different percentages of features are used in different models to see using how many features gives better performance. Supervised classification models are applied on the Dataset to predict the category of a crime incident.

V. THE PROPOSED MODEL

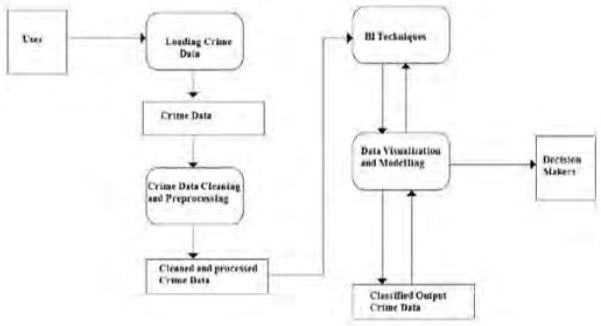
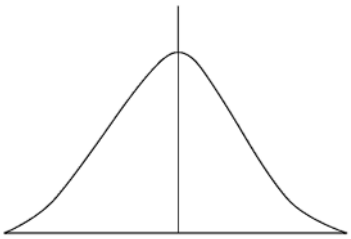


Fig.2 Flow of the process

The first step is to upload the dataset and then pre-process and then clean the data so that it is ready for the analysis. Noisy data will adversely have an effect on the results of any data mining analysis therefore it's of essence to clean up the data.The data flow diagram illustrates how data flows from process to process.The performance of different classification models is calculated and the best one is then selected.

1. Decision Tree: Sklearn.tree module provides DecisionTreeClassifier class. Among many parameters of this class, two parameters are useful in this case: min_samples_split indicates the number of splits to make at each step of building a decision tree and criterion indicates the function to measure the quality of split. As discussed above, there are two types of function to measure quality of split: information gain and impurity. In this class information gain is indicated with entropy and impurity is indicated with gini.

2. Gaussian Naive Bayes: In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a Gaussian



distribution. A Gaussian distribution is also called Normal distribution. When plotted, it gives a bell shaped curve which is symmetric about the mean of the feature values as shown below:

Fig.3 Gaussian Naive Bayes Graph

The likelihood of the features is assumed to be Gaussian, hence, conditional probability is given by:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

Sklearn.naive_bayes provides GaussianNB class.

3. KNearestNeighbors: The *k*-nearest neighbors algorithm (*k*-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the *k* closest training examples in the feature space. The output depends on whether *k*-NN is used for classification or regression: In *k-NN classification*, the output is a class membership. In *k-NN regression*, the output is the property value for the object. *k*-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. Both for classification and regression, a useful technique can be to assign weights to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. A peculiarity of the *k*-NN algorithm is that it is sensitive to the local structure of the data. KNearestNeighbors class in sklearn.neighbors module provides supervised nearest neighbors classification models using k nearest neighbors. Among different parameters of the class, n_neighbors indicates the value of k, metric indicates the metric used to measure the distance of neighbors. An equivalent of Euclidean distance is used as default.

4. Logistic Regression Classifier: Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative distribution function of logistic distribution. Thus, it treats the same set of problems as probit regression using similar techniques, with the latter using a cumulative normal distribution curve instead. Equivalently, in the latent variable interpretations of these two methods, the first assumes a standard logistic distribution of errors and the second a standard normal distribution of errors. Logistic regression can be seen as a special case of the generalized linear model and thus analogous to linear regression. sklearn.linear _model.LogisticRegression class is used for this model. The parameter multi_class is set to ovr which provides one vs the rest scheme,as a multi class model is needed. Class_weight parameter makes classes balanced in case of imbalanced classes.

5. Random Forest: Parameters of RandomForestClassifier class that are used are: n_estimator, min_samples_split, criterion, where min_samples_split and criterion are the parameters of decision tree of the random forest. N_estimator indicates the number of trees to build. There are a lot of benefits to using Random Forest, but one of the main advantages is that it reduces the risk of overfitting and the required training time. Additionally, it offers a high level of accuracy. Random Forest runs efficiently in large databases and produces highly accurate predictions by estimating missing data.

Our methodology consists of steps, data collection, data pre-processing, classification model building using training data and model evaluation using test data. It then uses a trained and tested model to score incoming data.

Finding crime patterns can be a very difficult task as we have to scan through a large amount of data, look for a pattern and then find if a particular crime fits the pattern. This process is very time consuming. If it doesn't fit, then the information should be classified as a new pattern.These reasonable results are going to be obtained once underneath going varied processes that return under machine learning.The first step would be to appear for an applicable information set for the system supported that the algorithms can offer India the output. These information sets are going to be correct data sets that are extracted from actual crimes that have taken place everywhere on the planet. A number of this information is going to be utilized in the coaching set and also the remaining will be used in the take a look at set.After this, we do the preprocessing of data. Here we can delete unwanted rows and convert improper information. For example we convert infinity to a numerical value or string into numerical.

The dataset that is considered for this project has been obtained from the official portal of the government of India. It consists of crime data related to all the states of India and also their respective districts. The collected data ranges across several years since having a good amount of historic data would help train the machine learning model to predict future crimes. There are several parameters in the data set like the name of the state, district, year, different types of crimes (Murder, Rape, Theft etc), total count. The dataset can be cleaned and processed as per the requirements of the project.

States/UTs	District	Year	Murder	Attempt to Culpable H	Attempt to Rape		Custodial	Custodial	Custodial	Rape othe	Rape_Gan	Rape_Oth	Attempt to	Kidnapping Ke
Andhra Pradesh	Anantapur	2014	134	171	8	0	35	0	0	0	35	0	35	1	125
Andhra Pradesh	Chittoor	2014	84	170	2	0	32	0	0	0	32	1	31	0	38
Andhra Pradesh	Cuddapah	2014	80	162	1	0	28	0	0	0	28	0	28	4	27
Andhra Pradesh	East Godavari	2014	64	84	2	0	85	0	0	0	85	0	85	18	66
Andhra Pradesh	Guntakal Railway	2014	14	4	0	0	0	0	0	0	0	0	0	0	0
Andhra Pradesh	Guntur	2014	105	137	4	0	49	0	0	0	49	0	49	24	64
Andhra Pradesh	Guntur Urban	2014	51	65	0	0	40	0	0	0	40	1	39	8	84
Andhra Pradesh	Krishna	2014	51	47	1	0	80	0	0	0	80	1	79	20	41
Andhra Pradesh	Kurnool	2014	118	135	5	0	32	0	0	0	32	1	31	4	53
Andhra Pradesh	Nellore	2014	78	117	4	0	58	0	0	0	58	3	55	16	158
Andhra Pradesh	Prakasham	2014	75	68	7	0	50	0	0	0	50	2	48	8	50
Andhra Pradesh	Rajahmundry	2014	18	18	1	0	55	0	0	0	55	1	54	3	15
Andhra Pradesh	Srikakulam	2014	30	29	4	0	40	0	0	0	40	1	39	3	33
Andhra Pradesh	Tirupathi Urban	2014	35	72	2	0	18	0	0	0	18	1	17	5	24
Andhra Pradesh	Vijayawada City	2014	23	45	1	1	64	0	0	0	64	0	64	4	42
Andhra Pradesh	Vijayawada Railway	2014	1	1	1	0	0	0	0	0	0	0	0	0	3
Andhra Pradesh	Visakha Rural	2014	43	50	4	0	38	0	0	0	38	0	38	0	24
Andhra Pradesh	Visakhapatnam	2014	38	57	0	0	84	0	0	0	84	0	84	0	130
Andhra Pradesh	Vizianagaram	2014	43	23	1	0	47	0	0	0	47	1	46	1	27
Andhra Pradesh	West Godavari	2014	90	85	4	0	146	0	0	0	146	1	145	46	67
Andhra Pradesh	Total	2014	1175	1540	52	1	961	0	0	0	961	14	947	185	1066
Arunachal Pradesh	Anjaw	2014	0	0	0	0	0	0	0	0	0	0	0	0	1
Arunachal Pradesh	Changlang	2014	9	3	0	0	6	0	0	0	6	0	6	0	11
Arunachal Pradesh	Crime Branch	2014	0	0	0	0	0	0	0	0	0	0	0	0	0
Arunachal Pradesh	Dibang Valley	2014	0	0	0	0	1	0	0	0	1	0	1	0	0
Arunachal Pradesh	Kameng East	2014	4	1	0	0	1	0	0	0	1	0	1	0	7
Arunachal Pradesh	Kameng West	2014	7	1	0	0	0	0	0	0	0	0	0	1	3
Arunachal Pradesh	Kaikang Kurrey	2014	4	1	0	0	3	1	0	1	2	0	2	0	2
Arunachal Pradesh	Lohit	2014	10	4	0	0	7	0	0	0	7	0	7	0	25

Fig.4 Crime Dataset of India

After this,we represent the information with the help of a diagram called a heat map. Dark colors are used to show more activity of crime and light shades are used to show less occurrence of crime.The figures explains how it happens. After seeing where the most crime is occuring, proper preventive measures can be taken into account. Increasing the police patrolling, installing more cameras and installing burglar alarms help notifying the authorities faster and help in preventing crimes.Dangerous time periods of various locations can also be highlighted based on the result obtained.

Fig.5 Graphical Visualization of Crime prone regions

Fig. 5 shows the regions that have a high probability for crime occurrence. The advantages of using heat maps over other representational mechanisms are:

- Numeric and category based color images.
- Gradient color range.
- Analyze only the data we want.
- Out of range data is automatically discarded.

So by knowing about the probable regions we can prevent crimes by taking preventive mechanisms like night patrolling, fixing burglar alarms, fixing CCTV cameras etc.

VI. CONCLUSION

One of the biggest problems that all law-enforcement and intelligence- gathering organizations are facing is accurately and efficiently analyzing the growing volumes of crime data. With the advancement of technology, sophisticated data mining and artificial intelligence tools are becoming easily available to the law enforcement community. These techniques can help process thousands of instructions in seconds, saving precious time. In addition, installing and running software often costs less than hiring and training personnel. Computers are also less prone to errors than human investigators, especially those who work long hours.

The main focus through this research work is to develop a tool that uses different data mining techniques to help law enforcement officers carry out their task with maximum efficiency, accuracy and speed. The proposed tool will clean the data set, identify crime patterns and analyse it along with classifying it into various suitable class labels with minimal expenditure. With the assistance of AI innovation, it has gotten simple to discover the connection and examples among different information. The work in this venture essentially rotates around foreseeing the kind of wrongdoing which may occur and the area of where it has happened. Utilizing the idea of AI we will fabricate a model utilizing training dataset that will experience information cleaning and information change.

The information will be pictured utilizing diagrams like bar, pie, line and dissipate charts and so on, each having its own attributes. Diagrams will help in understanding violations datasets that can help in catching the components that can help in protecting society. In this project, using Machine Learning Technology, various algorithms will be used to :-

A) Breaking down crime patterns
B) Foresee future spots for crimes
C) Determine dangerous time periods of various regions
D) Predict the type of crimes that could happen in different regions based on historic data and patterns
E) Decide techniques to prevent those crimes.

Different data sets that have been extricated from genuine cases can be utilized to prepare the framework. The informational indexes can be isolated into preparing sets and testing sets.
The testing informational collections will be utilized to test the comprehension of the machine and the exactness of the calculations to do the counteractive action, examination and expectation of crimes.
For this research, only crime data has been used, but as many researches have shown that a particular area's socio-economic standard is also a key indicator of possible criminal activity. This machine learning agent could incorporate that data and might perform better.

This model can be also used for other geographic locations. This would also help to analyze crimes occurring in different locations and build a better understanding of different crimes and its relation with particular demography. Also, there are many advanced machine learning approaches that can be explored.
Deep Learning & Neural Networks can provide a more balanced understanding of criminal activities. As it has been seen on this research, imbalanced classes has been a major issue in dealing with the particular database. Advanced techniques to deal with imbalanced classes are also something that remains to be explored.

References

[1] Malathi. A and Dr. S. Santhosh Baboo, "An Enhanced Algorithm to Predict a Future Crime using Data Mining" : International Journal of Computer Applications, May 2011

[2] Shiju Sathyadevan, Devan M. ,"Crime Analysis and Prediction Using Data Mining.Conference": First International Conference on Networks & Soft Computing (ICNSC), 2014

[3] Alkesh Bharati, Dr Saravana Guru, "Crime Prediction and Analysis Using Machine Learning" : RA.K International Research Journal of Engineering and Technology (IRJET), 2014

[4] Suhong Kim , Param Joshi, Parminder Singh Kalsi, and Pooya Taheri, "Crime Analysis Through Machine Learning" : 2011

[5] Lawrence McClendon and Natarajan Meghanathan, "Using Machine Learning Algorithms to Analyse Crime Data" : March 2015

[6] Ewart BW, Oatley GC (2003) Applying the concept of revic-timization: using burglars' behaviour to predict houses at risk of future victimization. Int J Police Sci Manag 5:69–84

[7] Redmond M, Baveja A., "A Data-driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments", European Journal of Operational Research, Science Direct, vol. 141, no. 3, pp. 660–678, 2002.

[8] S. Sathyadevan and S. Gangadharan. "Crime Analysis and Prediction Using Data Mining," Proceedings of the 1st International Conference on Networks and Soft Computing, pp. 406-412, 2014.Al Boni, M., & Gerber, M. S. (2016, December). Area-

[9] Specific Crime Prediction Models. In Machine Learning and Applications (ICMLA), 2016 15th IEEE

International Conference on (pp. 671-676).IEEE.

[10] Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and AerospaceTechnology (ICECA), 2017 International conference of (Vol. 1, pp. 225- 230). IEEE.

Smart Bag

Rishabh Chorasiya, Bobby Kalal
Thakur College of Engineering and Technology, Mumbai, India
rishabh.chaurasiya1@gmail.com, bobbykalal07@gmail.com

Abstract— **Smart Bag uses Internet of Things and provides solution to numerous things which includes giving an alert about the items left outside the luggage, allowing it to track its location if it goes missing. The concept of smart luggage was implemented to give the passengers to stop worrying about their luggage and rather enjoy their journey. It consists of various features such as remembering the belongings in the bag, tracking the location of the bag. These operations are carried out using mobile application.**

Keywords—Raspberry Pi, GPS, GSM, RFID Reader.

I. INTRODUCTION

Travelling has always been an important aspect of our lives. We travel to meet new people, have new and exciting experiences and explore the beautiful places that exist in this world. And to travel, everyone needs their stuff, i.e. their luggage, with them. But, in India, one of the biggest problems while travelling remains the theft of our luggage/bag. This has been a problem since long. Hence there is a need to keep our dear ones and our belongings safe. But no matter how much we try and no matter how many locks we put on our luggage, what would happen if this luggage gets lost or robbed in the middle of your journey? It would be a menace because it would contain all your important things or sometimes documents. Hence, to provide a solution to this problem would include introducing the concept of smart bags thereby providing a better security. It is one of the alternative solutions where instead of looking after your baggage every now and then, where the passengers follow the bag, the bag follows them. Also, if the bag is stolen, the passenger can always get its location with the GPS tracker in it. We have also tried to add some more features for better over-all user experience, such as time-table reminder, finger-print lock for better security, and some more features that will be useful in the future.

II. BACKGROUND

The Smart Bag using IOT provides solution to numerous things which include right from giving an alert about the items if left outside the luggage to allowing it to track its location if in case went missing with the help of the RFID Tags and Reader. The concept of smart luggage was implemented to give the passengers a relief from worrying about their luggage but to worry about their journey. It is used in tracking the location of itself in case if it went missing or if forgotten somewhere which would include the concept of GPS. Also, in case if their luggage got taken away by someone it provides us a facility for the locking of the bag immediately if required. This concept was further thought of narrowing to small bags of schools which would include the basic feature of just tracking the location for the children if required.

III. OBJECTIVE

The objective of this project is to provide security of the luggage inside the bag. The bag must be able to track and also provide alerts of the bag in real-time, if the bag is stolen or misplaced or forgotten by the owner. It should provide power backup for mobile devices and laptops for emergencies. It should also alert the owner if any book is missing from the scheduled time-table. Also, it should provide security to school-going children with security features, such as a panic button, which, when pressed, will send an alert to their parents.

IV. PROBLEM DEFINITION

The main objective of this project is to provide security and tracking of the luggage bag as well as the belongings of the bag owner through IoT. Also, it should provide power backup for mobiles and laptops. It should provide security to school-going children with security feature and alert their parents through message. The phases in which a part of problem will be solved is provided below:

- **Planning:** The project planning was done based on the need for the smart bag in the real world thereby giving importance to overcoming the drawbacks of it. This concept was selected by taking into consideration the actual need for it in daily travelling. The various features like location tracking, monitoring forgotten items, automatic lock and other features were planned and documented.
- **Analysis:** The analysis of the bag was done based on comparison with the features of other bags, which included the tracking of the location of the bag. By connecting the bag with the application enabled the users to automatically lock their bags just by providing them a minimum location. This location helped the user to realize that where is their bag and is the location of the bag moved ahead or the coordinates of the location changed which indicated the bag being taken somewhere. Also, analyzing helps understand various issues.
- **Design:** Construction of the bag according to the required specification is done in this phase. The bag is

made up of different hardware support. Also, there is a complete software support for the working of the bag. The hardware is connected to and used by the

software thereby making it easy to handle and use. The designing of the entire application and the required hardware setup is done in this phase. The block diagram for the hardware set up is done in this phase. The construction of the diagram helped in determining the placement of the components in the bag without affecting the overall usage of the bag. The application interface and the requirement specification along with the feature for the alert is designed in this phase.

- **Coding:** Writing the whole system app code and taking help of open source. The entire code for the application would be written in this phase and would be tested. Application development is can be done via different platforms available. The application thus developed should be user friendly and easy to use. The complete guidelines for its usage can be displayed at the beginning so as ease the customer for further usage. The whole hardware is connected to the bag which included of GPS, lock, the weighing component and other. An application was later built which would help in connecting to this hardware. The control of the hardware was given to this application. The application entire code was written using the basic languages. The hardware of the bag was setup as such not to increase the weight of the bag so as make it heavy. Also, the application was made user friendly as so the user was able to access it with ease.
- **Implementation:** Give the user for alpha testing and gathering centric analysis of performance, feedback and try to improve the quality of the result/output. In this phase the testing would be done for checking the hardware and the software support. Also feedback would be taken from customers and improvements would be made further if required. The testing of the bag was done which included checking and weighing the items of the bag and verifying whether it gave the desired results or not. The location tracking with automatic lock was tested by taking the bag to different locations and verifying whether the application was able to give an alert or not. The feedback from various customers was taken to improve the graphical user interface of the application.

V. ANALYSIS

The smart bag will help the owner be carefree of their luggage and focus on the journey and their work. It will also help the students to keep in check with their books and study better. Also, as they will only carry the required books, the burden on their shoulders will be less. Children and women security is a very critical feature in current date and time. It will also help the police authorities track the bag more efficiently and quickly, thereby, reducing their time and energy on such cases.

VI. PROPOSED SYSTEM

The proposed system uses the technique of Internet of Things in order to track the bags. In this, hardware would be created and installed which would be having the basic arduino board with a GPS module and an alarm being connected to it. A map has been created which would be synchronized in order to track the location of the bag. Furthermore, the map has the features that as soon as the bag gets lost or theft and it moves away from the owner and goes out of a particular range, the alarm would start ringing so that the owner gets notified where exactly the bag is. Also it would help the owner to track down the location of the bag which could be seen on the map as the markers would be dropped which gives us the location of the bag as it moves away from the owner.

VII. METHODOLOGY

The methodology that we chose for our project is Agile. While waterfall model is a tried and tested framework, it is not suitable for our project, due to the changeable nature and small team size. Agile allows us to be flexible, manage changing requirements, manage the ever-increasing scope as well as get consumer perspective. Agile has focus on customer satisfaction, which is a crucial aspect of any healthcare application. Patients need to be clued in to the process for the indispensable inputs they can provide.

Agile Methodology

Agile methodology mainly focuses on the incremental and the iterative development of the different project phases. It controls and determines how the various stages are implemented and what kind of environment is involved. The project is first divided into different phases, and then the responsibility of developing these phases is given in the hands of the various developers. At every stage of the project development it is continuously monitored and taken care of. These developers developing the phases have particular amount of time period and tasks involved with them. To remain in sync with the other members of the team they have to complete their task within the stipulated amount of time. Agile is considered to be a methodology that is ready to accept and work on any change that occurs during any phase of the project which can be from any side of the project. It is flexible methodology that welcomes change during any stage of the project. After the development phase each task or each developed phase is tested in different environment. This helps the team members to determine what would be the next phase of the project. Every iteration of the project involves a testing phase. It requires close communication with the project master along with team members and the product owner for the development of the project.

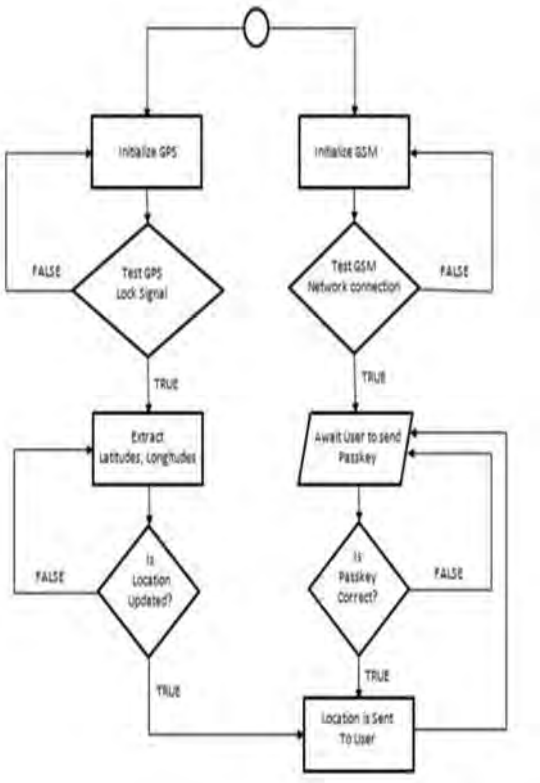


Fig 1: Methodology Flowchart of GPS of Smart Bag

VIII. FEASIBILITY STUDY

1. Executive Summary

Project is related to Smart Bags.

The project maintains two levels of users: -

- Administrator Level-Owner
- User Level-Data Entry Operator
- Main facilities available in this project are: -
- Maintaining the daily time-table.
- Providing GPS location tracking.
- Reminding customers of the luggage.
- Providing the weight capacity of the bag.
- User can set a reminder for his/her different requirements.

2. Define business problem or opportunity

The Software is for the automation of Smart Bags. It maintains one level of user:-

- User Level

The Software includes:-

- Maintaining customer details.
- Providing and maintaining all kinds of storage facilities for a customer.
- Reminding and giving an alert to the user as per requirements.

3. Requirements and purpose of the study

- Hardware Requirements
  - Raspberry pi
  - Bag
- Software Requirements
- Android Studio
- **Purpose**
- Delightful Bags provide following facilities like:-
- Taking orders from the customer.
- Delivering ordered meal timely
- Bill computation
- Record maintenance.

4. Description of the options assessed

Based on research conducted over the Internet we identified the following alternatives for the MMS:

- Continue to use the current system.
- Utilize a packaged solution. The bag providers who require this software can purchase the product and host them in their own environment. Less amount of programming and re- programming is required for customizing the system and implementing them.
- Develop a customized bag solution which meets all requirements satisfactorily and is also built with a scalable scope.

5. Assumptions used in the study

Every user should be:

- Comfortable bag usage.
- He/she must have proper internet connection.
- He/she must also have basic knowledge of English too.

6. Audience impacted

It can be used in anywhere as far the internet services are provided.

7. Financial obligation

The financial and the economic questions during the preliminary investigation are Verified to estimate the following:

- The cost to conduct a full system investigation.
- The cost of hardware and software for the class of application being considered.
- The benefits in the form of reduced cost.

The proposed system will initially give the minute information, as a result the performance is improved which in turn may be expected to provide increased profits.

IX. HARDWARE DESCRIPTION

The various components required for the development of the project Smart Bag using IOT along with their estimated budget is as shown below:

o RASPBERRY PI 4

o GPS module Ublox NEO 6mM GPS module

o SD card 64 GB Sandisk Ultra memory card

o Universal card Reader phone-PC smart card reader

o Bread Board 840 points solder less

o Female to Female Jumper wires

o Male to Male Jumper wires

o Berry GPS-GSM

o Passive RFID Tags and Reader

o Arduino Mega 2560 R3

o GSM module for SMS service

o Finger Print Scanner Module (R305)/ Fingerprint Scanner GT(511C1R)

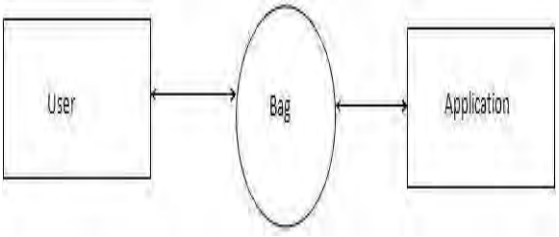
o Buzzer and Panic Button

o Douself Sola magnetic lock.

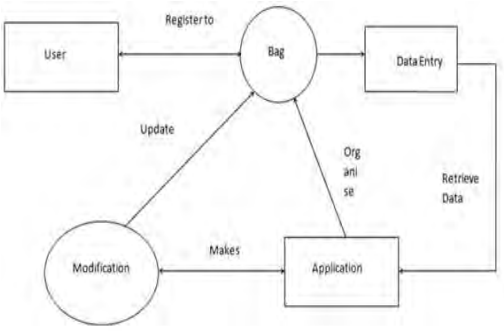
X. DESIGN AND IMPLEMENTATION

1. Data Flow Diagram (DFD)

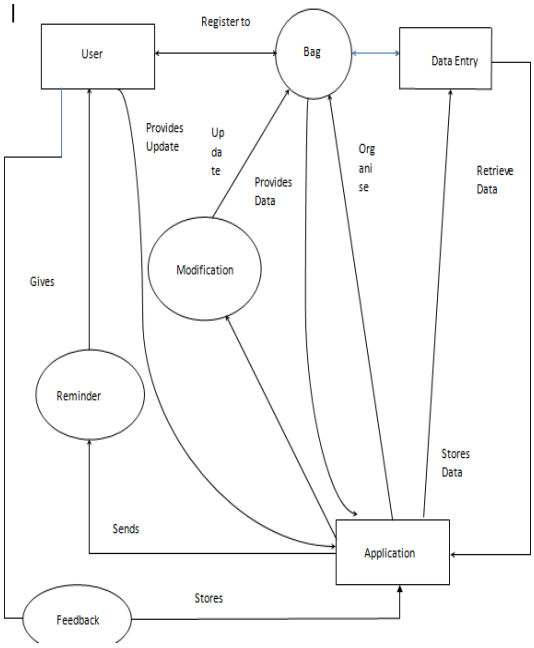
DFD level 0:



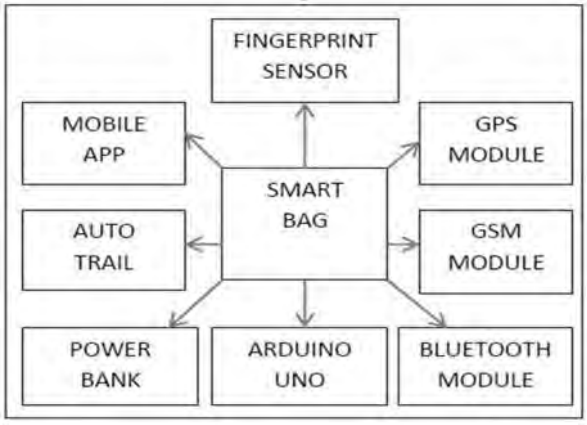
DFD level 1:

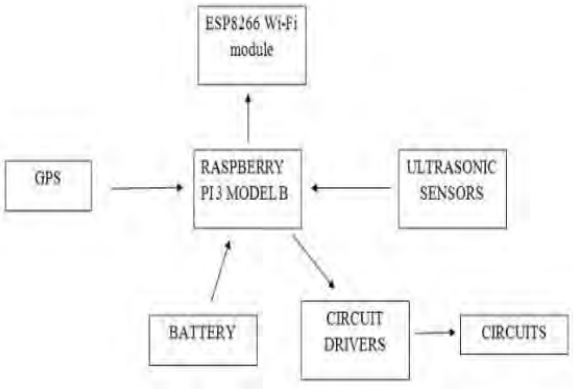


DFD level 2:



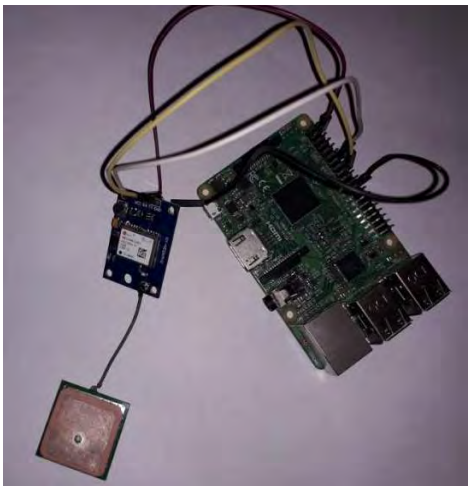
2. Block Diagrams





Implementation

This project idea shows the implementation of tracking the bags which are either lost or stolen using IoT. Certain procedures and techniques have been made and proposed in order to achieve the same. Experimentation has been done and maps have been created in order to track the location of the bags which are misplaced and lost. Experiment results further concludes that the bags can be easily tracked based upon the hardware installed in them and then by tracking that hardware and tracking the route, directions and location of the bag with the help of a map.



XI. RESULT

This smart bag consist of inbuilt circuit which provide location tracking and baggage remembering system to individual at cheapest cost. Additional feature has been added to facilitate Students not to miss textbook as per college/school time tables concept of Google Home mini and list of baggage (for backup) has been used to make sure all books are carried in the bag if any missing book from the time table will be notified.

XII. CONCLUSION

The progress in science & technology is a non-stop process. New things and new technology are being invented. As the technology grows day by day, we can imagine about the future in which thing we may occupy every place. The proposed system based on microcontroller is found to be more compact, user friendly and less complex, which can readily be used in order to perform several tedious and repetitive tasks. Though it is designed keeping in mind about the need for school kids, it can extended for other purposes such as commercial & research applications. Due to the probability of high technology (Android) used this system is well software controlled. The feature makes this system is the base for future systems.

Smart bag is an application-specific design that can be useful for almost everyone in the society. Smart means intelligent, the bag will be intelligent enough to carry out various features for everyday use .

XIII. FUTURE SCOPE

In future, we are planning to include some interesting features like automatic object avoidance, stair case climbing and some extra woman safety features. These extra features make the bag more powerful and user friendly. In future we can use this idea of smart bag to packing system in factories, shops, super markets etc.

References

[1] Rakhi Varma, Pooja Pavshe, Akshay Bhadane, Shrushti Pagare, “MULTIFUNCTIONAL BAG MONITORING SYSTEM”, 2018.

[2] Sebin J Olickal, Amal Yohannan, Manu Ajayan, Anjana Alias,” SMART BAG (It can follow you)”, 2017.

[3] Sudha Senthilkumar, Brindha.K, Rathi.R, Charanya. R, Mayank Jain,” LUGGAGE TRACKING SYSTEM USING IOT”, 2017.

[4] Shubham Sarkar, Suvojit Manna, Subhadeep Datta,” Smart bag tracking and alert system using RFID”, 2017.

[5] Mrs.Rasika Naik, Sanjana Muppidwar, Pallavi Chavan, Siddhi Medhekar, Pooja Chindarkar,” SMART BAG”, 2016.

[6] Sai Bhargavi.B, Praneetha, Siva Kumar D, Mohana Rao Chakka,” FINGERPRINT BASED DOOR ACCESSING SYSTEM USING RASPBERRY PI KIT”, 2017.

[7] Jordi Sapes and Francesc Solsona,” Finger-Print Scanner: Embedding a Fingerprint Scanner in a Raspberry Pi”, 2016.

[8] https://ieeexplore.ieee.org/document/8284576

[9] https://ieeexplore.ieee.org/document/6142188

[10] https://ieeexplore.ieee.org/document/8300839

[11] https://ieeexplore.ieee.org/document/7949485

[12] http://www.jetir.org/papers/JETIR1602017.pdf

Image Captioning Using Deep Learning

Rohan Arora, Meghanshu Bhatt, Akhil Daphara, Shridhar Kamble
Department of Information Technology,
Thakur College of Engineering and Technology, Mumbai,
rohanarora3434@gmail.com, meghanshubhatt@gmail.com, akhildaphara@gmail.com, shridhar.kamble@thakureducation.org

Abstract—**Image captioning, which aims to automatically generate a sentence description for an image, has attracted much research attention in cognitive computing. The task is rather challenging, since it requires cognitively combining the techniques from both computer vision and natural language processing domains. Existing CNN-RNN framework-based methods suffer from two main problems: in the training phase, all the words of captions are treated equally without considering the importance of different words; in the caption generation phase, the semantic objects or scenes might be misrecognized. In our work, we propose a method based on the encoder-decoder framework, named Reference based Long Short Term Memory (R-LSTM), aiming to lead the model to generate a more descriptive sentence for the given image by introducing reference information. Specifically, we assign different weights to the words according to the correlation between words and images during the training phase. We additionally maximize the consensus score between the captions generated by the captioning model and the reference information from the neighboring images of the target image, which can reduce the misrecognition problem. We have planned to conducted extensive experiments and comparisons on the benchmark datasets MS COCO and Flickr30k. By this approach we have try to outperform the state-of-the-art approaches on all metrics. We come to a conclusion that through the introduction of reference information, our model will learn the key information of images and generate more trivial and relevant words for images.**

I. INTRODUCTION

Benefiting from the significant advances of large-scale labelled datasets, such as ImageNet and deep learning, especially deep convolutional neural networks (CNN); the problems of image classification and object recognition have been studied thoroughly. As a result, computers even outperform humans at these tasks. Recently, automatically generating a sentence description for an image, has attracted much research attention in artificial intelligence. This problem, known as image captioning, plays an important role in computer vision, i.e., enabling computers to understand images, which can be exploited in wide applications, such as video tracking, cross-view retrieval, sentiment analysis, childhood education, and visual impairment rehabilitation. However, image captioning is a challenging task due to the coverage of both computer vision and natural language processing technologies. Apart from the need for identifying the objects contained in an image, the generator should also be able to analyze their states, understand the relationship among them, and express the semantic information in natural language. The early efforts on image captioning mainly adopt the template-based methods, which require recognizing the various elements, such as objects as well as their attributes and relationships in the first phase. These elements are then organized into sentences based on either templates or pre-defined language models, which normally end up with rigid and limited descriptions. As a typical transfer-based method, nearest neighbor (NN) is employed to retrieve a description from the corpus for a given image. Although this method cannot generate any novel sentence, it suggests that NN can indeed provide valuable information. Inspired by recent advances in machine translation, neural network-based methods have been widely applied in image captioning tasks and achieved great success. These methods are primarily based on the encoder-decoder pipeline, consisting of two basic steps. First, visual features are extracted using CNN to encode the image into a fixed length embedding vector. Second, recurrent neural network (RNN), especially long

short-term memory (LSTM) is adopted as the decoder to generate the sentence description by maximizing the likelihood of a sentence given the visual features. Thanks to the feature representation capability of CNN and the temporal modelling of RNN, the neural network- based methods are more flexible, which can generate new sentences coherently. On the other hand, motivated by the attention mechanisms, which have been proven to be effective in visual scene analysis, different attention mechanisms are proposed for image captioning, such as region-based attention, visual attention, semantic attention, global-local attention, and spatial and channel-wise attention.

A. Abbreviations and Acronyms

CNN: In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used.

LSTM: The LSTM (Long Short-Term Memory) layer is nothing but a specialized Recurrent Neural Network to process the sequence input (partial captions in our case).

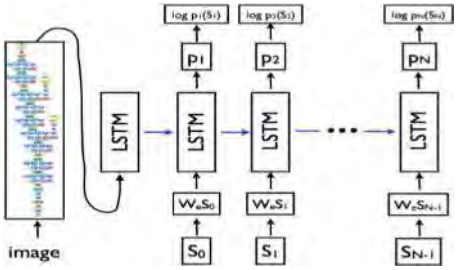
GRU: A gated recurrent unit (**GRU**) is part of a specific model of recurrent neural network that intends to use connections through a sequence of nodes to perform **machine learning** tasks associated with memory and clustering, for instance, in speech recognition.

II. Background

Generally, the existing image captioning algorithms can be divided into three categories based on the way of sentence generation: template-based methods, transfer-based methods, and neural network-based methods. The template-based methods either use templates or design a language model, which fill in slots of a template based on co-occurrence relations gained from the corpus, conditional random field, or web-scale n-gram data. The transfer-based methods are based on the retrieval approaches, which directly transfer the descriptions of the retrieved images to the query image. Some approaches took the input image as a query and selected a description in a joint image-sentence embedding space. The neural network-based methods come from the recent advantages in machine translation, with the use of RNN. A multimodal layer to connect a deep CNN for images and a deep RNN for sentences, allowing the model to generate the next word given the input word and the image. Inspired by the encoder-decoder model in machine translation, used a deep CNN to encode the image instead of a RNN for sentences, and then used LSTM, a more powerful RNN, to decode the image vector to a sentence. Many works follow this idea, and apply attention mechanisms in the encoder. With each feature representing a fixed-size region of the image, the model can learn to change the focusing locations. Employing a pre-trained CNN for object detection to analyse the hierarchically segmented image, and then ran attention-based decoder on these visual elements. Combining the whole image feature with the words obtained from the image by attribute detectors can also drive the attention model. Proposed a global-local attention mechanism by integrating local representation at object-level with global representation at image level. More recently, reinforcement learning has been integrated in the encoder-decoder framework. To address the deficiencies of exposure bias and a loss, which does not operate at the sequence level in traditional encoder-decoder frameworks, a novel sequence level training algorithm, named Mixed Incremental Cross Entropy Reinforce (MIXER), that directly optimizes the metric used at test time.

$$R = \arg\max_{R'} \mathcal{O}(R'|J;\theta^*), \quad (2)$$

III. IMPLEMENTATION DETAILS



Our goal is to generate a description sentence for an image. Suppose we have N training images I1, I2, ⋯ , IN , which also denote related visual features. For image In (n = 1, 2, ⋯, N), we have Mn correct description sentences Sn1, Sn2, ⋯, SnMn

Our task aims to maximize the likelihood of the correct descriptions given the training images by the following:

$$\theta^* = \arg\max_{\theta} \sum_{n=1}^{N} \sum_{m=1}^{M_n} \mathcal{L}(S_{\mathrm{nm}} | I_n; \theta), \qquad (1)$$

where θ are the parameters of our model and L() is a pre- defined likelihood function. In the next section, we will firstly describe the conventional likelihood function for image captioning used in previous works, and then we will introduce the proposed likelihood objective function.

After training, we can generate a sentence for a test image J by the following:

where O() is a pre-defined objective function. This objective function aims to generate the best sentence for the given image J.

Our model will have three main parts: a convolutional neural network (CNN) that extracts features from images, an attention mechanism that weights the image features, and an RNN that generates captions to describe the weighted image features. In the following sections, we will describe each in turn.

A convolutional neural network is a standard feed forward neural network, except that instead of computing the affine function f(x) = Ax + b, we restrict A to be a Toeplitz matrix. This is the same as saying that A is a convolution. This basic idea generalizes to images, where we have a 2D convolution over each of the K color channels. In the context of neural networks, we often follow these affine convolutional layers with a nonlinearity. In fact, f is an affine operator on x, and nonlinear itself. The most commonly used nonlinearity is the rectified linear unit (ReLU), which is nothing more than the hinge loss. While CNNs are adept at signal processing, they cannot easily express the idea of patterns in sequences.

Recurrent neural networks (RNNs) are similar to feedforward neural networks, except their outputs are fed back into the input. The networks maintain a hidden state that allows them to adjust their behavior throughout iterations of this feedback loop. RNNs are considered state of the art in machine translation and other tasks involving generating text, as they easily learn the patterns of human grammar. In this work, we leverage a powerful variant on the RNN idea called a Long ShortTerm Memory (LSTM) unit.

IV. RESULT

To understand how good the model is, let's try to generate captions on images from the test dataset (i.e. the images which the model did not see during the training)

No model in the world is ever perfect and this model also makes mistakes. Let's look at the example where the captions are not very relevant and sometimes even irrelevant.

As you can see the outputs, our model is not perfect but our objective is to further improve our model in future sprints by utilizing various deep learning techniques and to achieve the state of art.

A model can make lots of mistakes like wrongly identifying/classifying the object/entity/person, grammatical mistakes and not understanding the scenario of the given image. We must understand that the images used for testing must be semantically related to those used for training the model. For example, if we train our model on the images of cats, dogs, etc. we must not test it on images of airplanes, waterfalls, etc. This is an example where the distribution of the train and test sets will be very different and in such cases no Machine Learning model in the world will give good performance.

V. Future Scope

Of course, this is just a first-cut solution and a lot of modifications can be made to improve this solution like:

•Using a larger dataset.

•Changing the model architecture, e.g. include an attention module.

•Doing more hyper parameter tuning (learning rate, batch size, number of layers, number of units, dropout rate, batch normalization etc.).

•Use the cross validation set to understand overfitting.

•Using Beam Search instead of Greedy Search during Inference.

•Using BLEU Score to evaluate and measure the performance of the model.

•Writing the code in a proper object-oriented way so that it becomes easier for others to replicate.

VI. VI. CONCLUSION

We introduced a model that generates natural language descriptions of image regions based on weak labels in form of a dataset of images and sentences, and with very few hardcoded assumptions. Our approach features a novel ranking model that aligned parts of visual and language modalities through a common, multimodal embedding. We showed that this model provides state of the art performance on image-sentence ranking experiments. Second, we described a Multimodal Recurrent Neural Network architecture that generates descriptions of visual data. We evaluated its performance on both full frame and region-level experiments and showed that in both cases the Multimodal RNN outperforms retrieval

baselines. Although our results are encouraging, the Multimodal RNN model is subject to multiple limitations. First, the model can only generate a description of one input array of pixels at a fixed resolution. A more sensible approach might be to use multiple saccades around the image to identify all entities, their mutual interactions and wider context before generating a description. Additionally, the RNN receives the image information only through additive bias interactions, which are known to be less expressive than more complicated multiplicative interactions. Lastly, our approach consists of two separate models. Going directly from an image sentence dataset to region-level annotations as part of a single model trained end-to-end remains an open problem.

References

[1] Andrej Karpathy, Li Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions", INSPEC Accession Number: 15524180, 2015

[2] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", arXiv:1502.03044 [cs.LG], 2016

[3] Guiguang Ding, Minghai Chen, Sicheng Zhao, Hui Chen, Jungong Han, Qiang Liu

[4] Neural Image Caption Generation with Weighted Training and Reference", Print ISSN18669956, Online ISSN1866- 9964, 2018

[5] Zhihao Zhu, Zhan Xue, Zejian Yuan, "Image Captioning with Attention", arXiv:1807.03514 [cs.CV], 2016

[6] Image captioning with Keras. Medium- https://towardsdatascience.com/image-captioningwith-keras- teaching-computersto-describe-pictures- c88a46a311b8

[7] Vasanthi.G and Ramesh Babu.Y Department of ECE, DMI College of Engineering, Chennai, India. "Vision Based Assistive System for Label Detection with Voice Output" Jan-2014

[8] Bien, Z.; Park, K.-H.; Jung, J.-W.; and Do, J.-H. 2005. Intention reading is essential in human-friendly interfaces forthe elderly and the handicapped. IEEE Transactions on Industrial Electronics 52(6):1500–1505.

Intelligent Game Development

Aakash Jha, Sejal Chandak, Anjali Choudhary, Namdeo Badhe
Information Technology
Thakur College of Engineering and Technology Mumbai, India
aakashbjha@gmail.com, sejalchandak18@gmail.com, anjalichoudhary262@gmail.com, namdeo.badhe@thakureducation.org.c om

Abstract -**Machine Learning and Game Development have been correlated in a very compact manner since the second half of the 20th century. Historically, games have followed a specific script based on the ideology of the developer where various stages were predefined i.e. static in nature. Generally, games have always been comprised of a set of agents and objects that follow a certain set of paths determining cumulatively the state that the agent is in, in order to reach its goal defined by the developer without any learning component. Behavior of these agents and objects have always been based on simple rules and consecutively, the goals set for these agents and objects are very deterministic in nature. In modern times, game development fraternity has made great strides in integrating ML algorithms in order to make games more dynamic and ever evolving. The advent of ML algorithms has ushered in the inclusion of learning components. The proposed system aims to use ML algorithms to determine the behavior and decision-making process for NPCs (Non- player Character) and to train agents to adapt to the dynamic gaming environment.**

Keywords- ML (Machine Learning), NPC (Non-player Character) Artificial Intelligence (AI)

I. INTRODUCTION

Arcade was developed in the mid-20th century which was the first in a long line of games in the now ever-growing gaming industry. Gaming Industry has become one of the largest industries in terms of employment and revenue generation. From computers to consoles to mobile games to AR/VR technologies and to whatsoever the future might hold, the evolution of games has been directly proportional to the growth of the industry. Online Gaming has been a flagship breakthrough and a huge contributor to the gaming industry. MMORPG (Massively Multiplayer Online Role-Playing Games) is an example of the next generation of online gaming. ML is a subset of AI (Artificial Intelligence) that deals with teaching machines to self-train and grow. Algorithms are built by developers to self-improve by classifying patterns in the data provided.

ML is no stranger to Games and simulations there are many assets available to the Unity developer which are used to provide simulated machine intelligence. In modern times we do not need to fake or simulate intelligence rather, it can be created.

We have come far in the ML world; we can construct agents which learn from their environment to the extent that they beat their human builders. A complete system, for instance, may involve ML tools in-addition with more classic AIs like Behaviour Trees this is done to simulate a better, more unpredictable game. Most developers may use ML for gaming, it certainly helps game development in the following areas:

- Map/Level Generation
- Model Generation
- Artificial Players
- NPCs or Game AI

II. BACKGROUND

IN PAPER [1] which was a General Video Game AI for Evaluating Agents, Games and Content Generation Algorithms by Diego Perez-Liebana specifies the group of algorithms which is used for planning of the single-player using evolutionary algorithms (EA). The agents do not use the forward model to plan to execute the real game. Rather the game is played multiple times and in each play the game learns by itself and improvises in further turns.

IN PAPER [2] Deep Reinforcement Learning for General Video Game by Ruben Rodriguez Torrado published in 2018: In this paper the RL (Reinforcement Learning) agents get learning through trial and error interactions within an environment which is dynamic,

further in this paper the reward trade-off between long-term and short-term planning is balanced. Each and every algorithm mentioned in this paper is trained on one million frames of any particular game

IN PAPER [3] The Mastering Chess and Shogi by Self-Play paper published in 2017 by David Silver reuses the existing hyper-parameters for all games and does not actually specify the tuning of the game. A single Neural networks concept is used which updates continuously using Alpha Zero concept instead of waiting for the iteration to complete but neural networks failed to upstage the grandmaster in chess.

IN PAPER [4] The 2016 paper of Creating AI Characters for Fighting Games by Giovanna Mart, Richard Cant and David Woods states the Round Robin tournaments between the standard AI

characters and the general AI characters. This paper highlights that the evolutionary process was better than that of the hand coded ones. But the extensive experimentation was not much focused in it.

IN PAPER [5] Review of the Use of AI Techniques in Serious Games Maite Frutos- Pascual Deustotech published in 2015: This paper explains that the use of Fuzzy logic is based on approximate reasoning. It is greater than binary approximations, and was specifically designed in such a way so as to cope with areas that are gray i.e they go beyond true or false.

IN PAPER [6] Real-Time Monte Carlo Tree Search in Ms Pac-Man by Tom Pepels published in 2014: In this paper Selection, Expansion, Playout, Backpropagation are the four main steps of the algorithm which is explained in detail in the paper and applied on the MS Pac-Man game.

III. PROBLEM DEFINITION

Traditionally games were based on a fixed storyline and are very limited in terms of action and dynamic nature of the play. The proposed system ensures that the games developed with machine learning reacts and responds more dynamically and in more imaginative ways to the player.

The proposed system can be implemented in 2 phases as follows:

Phase 1: Develop a game using “unity beta” development environment.

Phase 2: Apply Machine Learning algorithms to the game that we had developed in the first phase in order to provide users with a better gaming experience.

IV. PROPOSED WORK

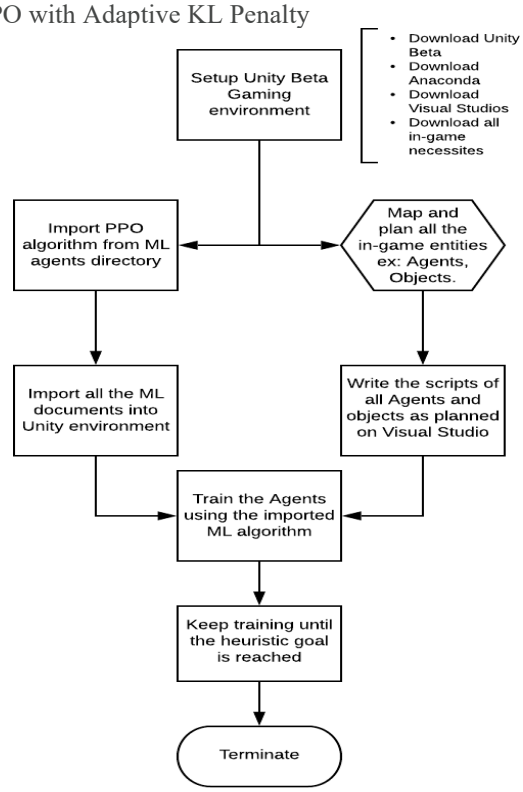
The proposed system uses unity ML agents to do ML and specifically reinforcement learning using Proximal Policy Optimization (PPO) algorithm to train agents that will try to learn and carry out certain actions based on the environment. PPO algorithm is extracted from the ML agents file which can be procured over the internet. This algorithm is used to train.

A Unity Beta game development platform is installed and studied. After the installation the ML agents are imported from the SDK assets directory. The scripts are written for the various agents that are required for the game using c- sharp. In the proposed system we made scripts for various agents and objects. Various 3D meshes of the agents and objects specifically required for the game are downloaded and these 3D meshes are imported from the asset’s directory. The meshes can be seen using the inspector view in unity. The positioning as well as the various other aspects of the agent, environment, object is designed according to the game. All the entities are placed in the area specified. Now an academy is created and the agent academy script is added to it. The agent learning brain is also added to the academy to broadcast the hub. In order to start the training a couple of configurations are setup. The trainer config DML is dragged in Visual Studio and the hyper parameters for machine leaning are set. A curriculum is set where the agent folder will have a text file called agent learning JSON which is an exact match to the name of the brain. Training of the agents can be done using Anaconda. The ML agent’s environment is setup on the Anaconda prompt and the agents are trained. Converting the brain type from player to heuristic in order to let the game learn by itself.

The PPO algorithm is an on-policy learning method which is used for navigating any game environment. The advantage of using the PPO algorithm is that it is compatible with the updates in the policies and thereby ensures less variance in training cost and also makes sure that the agent does not take any senseless actions.

Fig. Flowchart Of Project

PPO with Adaptive KL Penalty



Step 1. Input: initial policy parameters θ_0, initial KL penalty β_0, target KL divergence δ

Step 2. for k = 0,1, 2...do

Step 3. Collect set of partial trajectories Dk on policy $\Pi k = \Pi(\theta k)$

Step 4. Estimate advantages $\hat{A}$ $\pi k\tau$ using any advantage estimation algorithm

Step 5. Compute policy update

$\theta k+1 = \arg\max\theta \angle\theta\kappa (\theta) - \beta k \bar{D} KL (\theta||\theta k)$

By taking K steps of minibatch SGD(via adam)

Step 6. If $\bar{D}$ $KL(\theta k+1||\theta k)\geq 1.5\delta$ then

$\beta k+1=2\beta k$

else if $\bar{D}KL(\theta k+1||\theta k)\leq\delta/1.5$ then $\beta k+1=\beta k/2$

end if

Step 7. end for.

V. METHODOLOGY

This project aims at developing a ML trained game. The main advantage of using ML algorithms in game development is that it learns the goal-oriented moves on its own gets better and better at those moves the more it trains thereby providing an excellent and improvised gaming environment.

The process starts by setting up the gaming environment which includes downloading various tools namely Unity Beta, Anaconda, Python IDLE, Visual Studios and other in-game accessories. Making sure that all these tools and accessories have been correctly setup is very important to the project. After the setup process is complete the planning and visualization of the entire project needs to be accomplished by defining the entire game which encompasses the Environment, Features, Goal, Agents, Objects, etc, and also everything that is required to correctly define all these entities. Once the game is visualized a training or learning algorithm needs to be picked to be used in this game, for the purposes of this project PPO (Proximal Policy Optimization) algorithm is being used. The algorithm needs to be written in a python environment so that it can be used to train the agents in the game.

Once the algorithm is written, all the files made in the writing of the algorithm need to be imported in the Unity Beta gaming engine. To visualize all the planned Agents and Objects of the game a script must be written for each and every one of them. This script must be written in Visual Studio, it defines all the characteristics of the said agent or object. Similarly, the scripts of gameplay, area, environment must also be written. This will create an entire scene in the Unity Beta gaming environment. Once the scene is created the Agent is trained using the PPO algorithm. The training needs to be continued until the agent masters the game completely.

VI. TOOLS

Various tools used to implement the project:

1.Visual Studio 16.5.0: For writing scripts (in C sharp) in order to create the scene.

2.Anaconda 5.3.0: Used for importing python files into Unity.

3.Unity Beta 2019.1b: Cross platform gaming engine used to develop the game.

4.Python 3.6: Programming platform that is preferred in this project to write training algorithm.

VII. VII. RESULTS AND DISCUSSIONS

Initially while getting accustomed to unity engine environment and PPO algorithm, a small-scale game namely the racket game was developed in this project. In this game the racket agents where given a heuristic brain which used the PPO algorithm from the ml-agents directory to train in accurately hitting the ball at all times.

The PPO learning algorithm is observed to have run very smoothly while making this racket game as this game was done on a small scale with single agent and object environment. It remains to be seen how this algorithm fares for our large-scale project including multiple agents and objects in the coming months of this project.

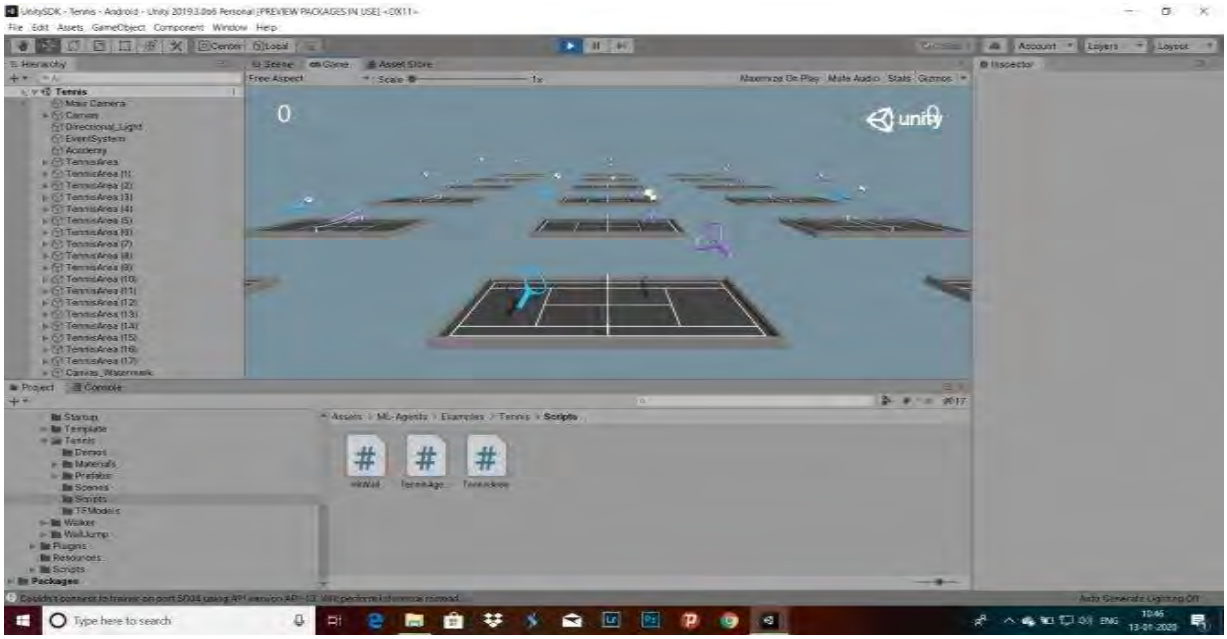


Fig. Racket Game

VIII. CONCLUSION AND FUTURE WORK

From the literature survey we observe that initial gaming models were based on storylines which made them static as there were fixed standard paths for each level. This made the games predictable and easy. In order to provide better gaming experience, the proposed system uses PPO (Proximal Policy Optimization) algorithm, method mapping, game stakeholder (end user) taxonomy and the player-game interaction loop. This analysis revealed dominant PPO (Machine Learning) algorithm within particular areas as well as in general areas has a room for exploration of new methods within the existing system. PPO algorithm is compatible with the updates in the policies and thereby ensures less variance in training cost and also makes sure that the agent does not take any senseless actions. But PPO algorithm finds it difficult to support GPU enabled implementations and to work on large scale projects and hence many updated versions of PPO namely PPO2 and ACER (Actor critic with experience replay) which claim to be 3 times faster than current PPO.

In future this project will deal with implementing a large-scale game with multiple agents and objects in an environment and training specific agents to play the game.

IX. ACKNOWLEDGMENTS

We sincerely thank our guide Mr. Namdeo Badhe for his guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Rajesh Bansode, the Principal, Dr.B. K. Mishra, and the college management for their support.

References

[1]. G. N. Yannakakis and J. Togelius, Artificial Intelligence and Games. Springer, 2018.
[2]. S. J. Russell and P. Norvig, Artificial Intelligence: a Modern Approach. Malaysia; Pearson Education Limited,, 2016.
[3]. D. Perez-Liebana, S. Samothrakis, J. Togelius,
[4]. M. Lucas, and T. Schaul, "General Video Game AI: Competition, Challenges and Opportunities," in Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 4335–4337.
[5]. M. Ebner, J. Levine, S. M. Lucas, T. Schaul, T.Thompson, and J. Togelius, "Towards a Video Game Description Language," Dagstuhl Follow- Ups, vol. 6, 2013.
[6]. T. Schaul, "A Video Game Description Language for Model-based or Interactive Learning," in IEEE Conference on Computational Intelligence in Games (CIG), 2013, pp. 1–8.
[7]. "An Extensible Description Language for Video Games," IEEE Transactions on Computational Intelligence and AI in Games, vol. 6, no. 4, pp. 325–331, 2014.
[8]. D. Perez, S. Samothrakis, J. Togelius, T. Schaul, S. Lucas, A. Couetoux, ¨ J. Lee, C.-U. Lim, and T. Thompson, "The 2014 General Video Game Playing Competition," IEEE Transactions on Computational Intelligence and AI in Games, vol. 8, pp. 229–243, 2015.
[9]. R. D. Gaina, D. Prez-Libana, and S. M. Lucas, "General Video Game for 2 Players: Framework and Competition," in 8th Computer Science and Electronic Engineering (CEEC). IEEE, Sep. 2016, pp. 186–191.

Air Pollution Prediction using SVM Algorithm

Vishal Bhat, Tushar Balande, Ruchi Bhavsar,
Information Technology Thakur College of Engineering and Technology
Mumbai, India
vishalbhat4444@gmail.com, balandetushar2@gmail.com, ruchibhavsar98@gmail.com

Abstract---**Air Pollution and its precaution has been a scientific as well an industrial challenge since past few decades. And it still remains an endless global problem. Thanks to the growing population and increase in settlements in urban areas, the quantity of pollutants in air increase day by day. Affecting human's respiratory and circulatory system, they are the main cause for increased mortality and increased risk for diseases for the population. Many efforts are taken by the governmental bodies to know and predict the AQI (Air Quality Index) to further improve public health. The foremost important step in prediction, no doubt, is to develop a predictive model of air quality status, which can help in management of the environment and also to make a way of awareness among people. Air quality prediction is the most important factor for monitoring pollution.**

Keywords---Air pollution, SVM, Machine Learning, Prediction.

I. INTRODUCTION

Artificial Intelligence and machine learning have seen a sudden rise within the past few years. Regarding the field of artificial intelligence where the machine makes all the choices on its own instead of having to program the machine for each specific need of the user, has started integrating itself in every aspect of our life. AI and machine learning have become the key focus of each company and organization, from start-ups to documented vendors.

The procedure of machine learning is that the machine first senses the environment and therefore the factors associated with it. It then gathers information and makes appropriate decisions like a human mind does naturally. One among the most important reasons why machine learning was chosen to predict air quality index and other harmful pollutants was this ability of adapting of machine learning (ML) algorithms.

The air is crammed with pollutants which are present in different states like liquid, gaseous or solid. These pollutants have a varying concentration, which

could determine its hazardousness. If the concentration of some pollutants is high, it could severely affect citizenry, plants, animals and environment. There are multiple sources which contribute to tremendous increase of air pollutants like indoor activities as well as outdoor activities. Many domestic activities like cooking, burning of bio-fuels and heating produce indoor air pollutants. There are two sorts of outdoor air pollutants namely, primary and secondary air pollutants. The pollutants that are released by industrial activities and fuel combustion into atmosphere are primary pollutants, whereas the secondary pollutants are a result of reaction in atmosphere. The varied harmful air pollutants are Oxides of nitrogen, particulate, carbon monoxide gas and CO2.

AQI Category (Range)	PM_{10} 24-hr	$PM_{2.5}$ 24-hr	NO_2 24-hr	O_3 8-hr	CO 8-hr (mg/m^3)	SO_2 24-hr
Good (0-50)	0-50	0-30	0-40	0-50	0-1.0	0-40
Satisfactory (51-100)	51-100	31-60	41-80	51-100	1.1-2.0	41-80
Moderate (101-200)	101-250	61-90	81-180	101-168	2.1- 10	81-380
Poor (201-300)	251-350	91-120	181-280	169-208	10.1-17	381-800
Very poor (301-400)	351-430	121-250	281-400	209-748*	17.1-34	801-1600
Severe (401-500)	430+	250+	400+	748+*	34+	1600+

II. BACKGROUND

In the last decade, artificial neural networks (ANNs) and more recently support vector regression (SVR) have emerged as two attractive tools for nonlinear modelling especially in situations where the event of phenomenological or conventional regression models becomes impractical or cumbersome. In recent years, support vector regression (SVR) which may be a statistical learning theory based machine learning formalism is gaining popularity over ANN, thanks to its many attractive features and promising empirical performance. The most important difference between conventional ANNs and support vector machines (SVM) lies within the risk minimization principle. Conventional ANNs implement the empirical risk minimization (ERM) principle to attenuate the error on the training data, while SVM adheres to the Structural Risk Minimization (SRM) principle seeking to line up a boundary of the generalization error. This study is motivated by a growing popularity of support vector machines (SVM) for regression problems. This process results in better generalization than conventional methods. This paper presents a study of using the SVM model to research of air pollutants. The SVM was trained by was performed on data of the principal air pollutants Nitrogen Dioxide, Sulphur Dioxide, Suspended PM, RSPM.

III. LITERATURE REVIEW:

1. Chi-Man Vong et. al.[1] suggested that to effectively compare accuracy of the model, four error measures namely, mean absolute error (MAE), root mean squared error (RMSE), complementary Willmott's index of agreement (CWIA), and relative error (RE) should be used. Predicted and observed level of NO2 in 1-year experiment, SPM in winter experiment, and O3 in summer experiment were plotted. Although there were some lagging and underestimations, the predicted levels produced by Linear model and RBF model followed the trend of observed level of NO2 pretty well in 1-year experiment. The prediction results of Linear model and RBF model showed a relative good fit to the observed test set of over one year of data, particularly for SO2 and NO2.

2. Soawalak Arampongsanuwat et. al. [2] studied and presented the potential of applying SVM model in ambient air pollutant prediction in this paper. The model was developed by using a technique from support vector regression (SVR). Gaussian Radial Basis Kernel functions was used because it is more suitable for the estimation of function Parameters to define the C (Regularization Parameter) with the ε (Error-insensitive) supported by vector machine. Parameters of Gaussian Radial Basis Kernel functions (σ) from the experiment find for the most appropriate using the values in the ascending range 0.1 through 1.0. In the experiment, the model selected has the lowest mean squared error (MSE). The results of this research show the model that set with value C = 5,000, ε = 0.001 and σ = 0.1 work out most precise in forecasting over other tested models. Based on the test forecasting data, the mean squared error (MSE) was 1.0588×10-10, which means this model is very satisfactory. The model reports that support vector regression can be used in forecasting PM10 successfully.

3. Shahid Ali et. al. [4] conducted an empirical study of SVM approaches to assess the capability of SVM in handling air pollution data set. A real-time dataset was obtained from USA and carried out rigorous experiments with single SVM, and ensemble methods like Bagging and AdaboostM1. From experimental results, it can be concluded that, ensemble methods outperformed single SVM in both accuracy and efficiency. Experimental results on air pollution dataset indicated that the proposed SVM with AdaboostM1 algorithm performs better than other algorithms. The classification accuracy of single SVM method was 76.33%t whereas with Bagging algorithm it was 79.66%. However, through AdaboostM1 algorithm, classification accuracy of 91.28% was achieved and less time was required to build ensemble model.

4. Bedoui, S. et. al. [6] Analysed the dispersion of ozone and the suspended particles presented. In this paper, Support Vector Machines SVM has been performed to predict the concentrations of the ozone and particles. To achieve this goal, the SVM with the linear kernel, SVM with the polynomial kernel and SVM with the RBF kernel have been used. The concentrations of pollutants have been investigated by a discriminant analysis DA of the monthly average values measured at four, and then at two locations in Tunisia. In the case of four classes, regardless of either linear or quadratic no reliable classification is available, with an error rate of 37.78%. However, in the case of two classes, the quadratic DA produces a lower misallocation rate (0%) than the linear DA (4.35%). A thematic map was created using the MapInfo software enabling classification of the regions.

5. Z. Ghaemi et. al. [7] designed a spatio-temporal system using a LaSVM-based online algorithm. Pollutant concentration, meteorological data and geographical parameters are fed to the online forecasting system. Performance evaluation of the system is done by comparing the prediction results of the Air Quality Index with those of a traditional SVM algorithm. Results show an increase of speed while preserving the accuracy of the SVM classifier. Comparison of the hourly predictions for next coming 24 h, with those of the measured pollution data in Tehran pollution monitoring stations shows an overall accuracy of 0.71, root mean square error of 0.54 and coefficient of determination of 0.81. These results are indicators of the practical usefulness of the online algorithm for real-time spatial and temporal prediction of the urban air quality.

IV. SVM WORKING

Support vector machines (SVMs) are known as an excellent tool for classification and regression problems, producing good generalization. The basic principle of SVM applies linear model to convert nonlinear class boundaries through some nonlinear mapping of the input vector into the high-dimensional feature space. The concept of a maximum margin hyperplane only applies to classification. However, support vector machine algorithms have been developed for numeric prediction that share many of the properties encountered in the classification case: they produce a model that can usually be expressed in terms of a few support vectors and can be applied to non-linear problems using kernel functions. Similar with linear regression, the basic idea here is to find a function that approximates the training points well by minimizing the prediction error. The crucial difference is that all deviations up to a user specified parameter are simply discarded. Also, when minimizing the error, the risk of over-fitting is reduced by simultaneously trying to maximize the flatness of the function. Another difference is that what is minimized is normally the predictions' absolute error instead of the squared error used in linear regression. A user-

specified parameter defines a tube around the regression function in which errors are ignored.

V. PROBLEM DEFINITION

Air Pollution is considered to be a concerning issue for the people and the environment. Therefore, many countries develop a monitoring system to reduce the effects caused by the pollution. Monitoring help us to examine the presence of noxious pollutants in the air, such as CO, NOx and NO_2. Due to the severe impact of air pollutants on the environment, monitoring and prediction of air has become important. The existing air quality monitoring stations in urban areas continuously record high volumes of pollutant concentrations. These data need to be effectively utilized for analysis and prediction of the air quality indices. The Support Vector Machine (SVM) model will be used to monitor the concentrations of gases in air and further predict values for the same.

VI. METHODOLOGY

- **Data Sampling**

The raw air pollutants and meteorological data from year 2005 to year 2010 were obtained from the website (https://www.kaggle.com/shrutibhargava94/india-air-quality-data).

state	location	type	so2	no2	rspm	spm	location	pm2_5	date
Maharashtra	Greater Mumba	Industrial Areas	34.8	31	43	293	Dombival	NA	05-04-2005
Maharashtra	Navi Mumbai	Industrial Areas	12.5	19.8	38	81	MIDC Taloj	NA	15-06-2006
Maharashtra	Navi Mumbai	Industrial Areas	12.7	47.7	28	124	MIDC Taloj	NA	09-04-2006
Maharashtra	Navi Mumbai	Industrial Areas	17.2	45.8	26	266	MIDC Taloj	NA	09-07-2006
Maharashtra	Navi Mumbai	Industrial Areas	13.7	28.7	21	82	MIDC Taloj	NA	09-11-2006
Maharashtra	Navi Mumbai	Industrial Areas	10.2	30.3	21	102	MIDC Taloj	NA	14-09-2006
Maharashtra	Navi Mumbai	Industrial Areas	8.8	26.7	30	130	MIDC Taloj	NA	18-09-2006
Maharashtra	Navi Mumbai	Industrial Areas	12	28.8	43	82	MIDC Taloj	NA	21-09-2006
Maharashtra	Navi Mumbai	Industrial Areas	7	12.5	53	143	MIDC Taloj	NA	25-09-2006
Maharashtra	Navi Mumbai	Industrial Areas	6.7	17.8	37	153	MIDC Taloj	NA	28-09-2006
Maharashtra	Navi Mumbai	Industrial Areas	22.5	40.3	124	557	MIDC Taloj	NA	12-04-2006
Maharashtra	Navi Mumbai	Industrial Areas	49.3	57.8	210	301	MIDC Taloj	NA	12-07-2006
Maharashtra	Navi Mumbai	Industrial Areas	121.8	50.2	26	163	MIDC Taloj	NA	12-11-2006
Maharashtra	Navi Mumbai	Industrial Areas	54	55.5	169	371	MIDC Taloj	NA	14-12-2006
Maharashtra	Navi Mumbai	Industrial Areas	60.8	42.7	143	386	MIDC Taloj	NA	18-12-2006
Maharashtra	Navi Mumbai	Industrial Areas	62	55.3	151	385	MIDC Taloj	NA	21-12-2006
Maharashtra	Navi Mumbai	Industrial Areas	51.3	57.7	147	373	MIDC Taloj	NA	28-12-2006

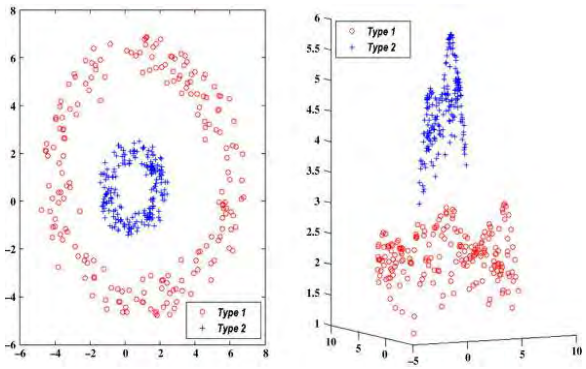
- **Data Normalization**

Prior to modelling, it is necessary to normalize all selected features into same range to avoid the domination by any feature with large values. This normalization process leads to more stable and accurate predicted results.

- **Modelling and Data Representation**

Air pollutants and meteorological data were used in this study to predict air pollutant levels. In order to apply SVM for pollutant level forecasting, the representation of a pollutant level is defined as a pair ($\mathbf{x}$, y).

The set of kernel functions is composed of variants in that they are all based on calculating inner products of two vectors. This means that if the functions can be recast into a higher dimensionality space by some potentially nonlinear feature mapping function, only inner products of the mapped inputs in the feature space need be determined without us needing to explicitly calculate ϕ. The reason that this kernel trick is useful is that there are many regression problems that are not linearly regressable in the space of the inputs, which might be in a higher dimensionality feature space given a suitable mapping function.



Then a data set that it not linearly separable in the two-dimensional data space is separable in the nonlinear feature space defined implicitly by this nonlinear kernel function known as radial basis function. Other popular kernels for classification and regression are the polynomial kernel.

In summary, in order to use an SVM to solve a regression problem on data that is not linearly separable, we need to first choose a kernel and relevant parameters which you expect might map the nonlinearly separable data into a feature space where it is linearly separable. This is more of an art than an exact science and can be achieved empirically, e.g. by trial and error. Sensible kernels to start with are the radial basis, polynomial and sigmoidal kernels. Therefore, the first step consists of choosing our kernel and hence the mapping function.

VII. CONCLUSION

Using observed meteorological and pollutant data, SVM models for forecasting ambient air pollutants were constructed. The prediction results of Linear model and *RBF model* showed a relative good fit to the observed test set of over one year of data, particularly for SO_2 and NO_2. It was evident that using Linear kernel and *RBF* kernel in SVM model for air pollutant forecasting produced superior results with

relatively lower errors. It is believed that SVM model with Linear kernel or RBF kernel can also produce good performance for air pollutant forecasting in other similar developing cities, or even other time series prediction in similar situation.

How to solve this problem to improve the accuracy of prediction model is the future works. Some literature attempted to integrate discrete wavelet transform (DWT) with SVM for a higher accuracy. Hence, we will attempt to integrate other machine learning methods, for example, genetic algorithm (GA), with SVM to improve the accuracy and efficiency of model in the future.

VIII. FUTURE SCOPE:

A much larger dataset can be used for better accuracy and prediction. Further, additional factors that might have affected the concentration of pollutants in a particular area can also be considered. More pollutants can be used as input to gain a great control over the pollution. Conventional methods such as the support vector machine (SVM) or artificial neural networks (ANN) show some inadequacy when enormous amount of data has to be analyzed for prediction. To overcome the limitations and improve the performance of prediction, a spatio-temporal system can be designed in the future. To predict the large dataset of pollutants provided by pollution control boards neural networks can be used. A better update into the SVM will result in advanced accuracy and efficiency.

References

[1] Chi-Man Vong, Weng-Fai Ip, Pak-kin Wong, and Jing-yi Yang, "Short-Term Prediction of Air Pollution in Macau Using Support Vector Machines" 2012.

[2] Soawalak Arampongsanuwat and Phayung Meesad,"Prediction of PM10 using Support Vector Regression" 2011.

[3]Jiang, D., Zhang, Y., Zeng, Y., Tan, J., Shao D., "Progress in developing an ANN model for air pollution index forecast" 2004.

[4] Shahid Ali and S S Tirumala, "Performance Analysis of SVM ensemble methods for Air Pollution Data" 2016.

[5] Ostro, B.D., Eskeland, G.S., Sanchez, J.M., Feyzioglu, T., "Air pollution and health effects: a study of medical visits among children" 1999.

[6] Bedoui, S.1, Gomri, S.2 , Samet, H.1 and Kachouri, A.1, "A prediction distribution of atmospheric pollutants using support vector machines" 2016.

[7] Z. Ghaemi, A. Alimohammadi & M. Farnaghi, "LaSVM-based big data learning system for dynamic prediction of air pollution" 2018. Global West International, 2000.

Stock Market Prediction

Neel Patel, Ahmed Tanwar, Harshita Khandelwal, Sudhir Dhekane
Information Technology
Thakur College of Engineering and Technology
Mumbai,India
neelpatel129@gmail.com, ahmedtanwar23@gmail.com, harshitaak17@gmail.com, sudhirdhekane1987@gmail.com

Abstract- **Stock market is basically nonlinear in nature and the research on stock market is one of the most important issues in recent years. To predict the stock market prices people search such methods and tools which will increase their profits, while minimize their risks. There are various algorithms for predicting stock market in Machine Learning. Due to various limitations like over-fitting and non-relevant inputs in Artificial Neural Network (ANN), algorithms like PSO (Particle Swamp Optimization) and LS-SVM (least square support vector machine) are gaining importance. This project aims to develop a machine learning model that hybrids the PSO and LS-SVM model. The performance of LS-SVM depends upon the selection of free parameters – C (cost penalty), ∈(insensitive loss function) and γ (kernel parameter). PSO will be used together to find the best parameter combination for LS-SVM.**

I. INTRODUCTION

Stock market plays a really vital role in quick economic process of the developing country like Asian country. Therefore our country and different developing nation's growth could depend upon performance of stock exchange. If stock exchange rises, then countries economic process would be high. If stock exchange falls, then countries economic process would be down. In different words, we are able to say that stock exchange and country growth is tightly delimited with the performance of stock exchange.

Predicting however the stock exchange can perform is one amongst the foremost tough things to try and do. There square measure such a large amount of factors concerned within the prediction – physical factors vs. physiological, rational and irrational behavior, etc. of these aspects mix to create share costs volatile and really tough to predict with a high degree of accuracy. Broadly, stock exchange analysis is

split into 2 elements – basic Analysis and Technical Analysis, however our focus is on the technical analysis half. Fundamental Analysis involves analyzing the company's future profitableness on the premise of its current business setting and money performance. Technical Analysis, on the opposite hand, includes reading the charts and mistreatment applied math figures to spot the trends within the stock exchange. Artificial Neural networks (ANNs) is the most commonly used technique. In most cases ANNs suffer from over-fitting problem due to the large number of parameters to fix, and the little prior user knowledge about the relevance of the inputs in the analyzed problem.

Also, Support vector machines (SVMs) had been developed as an alternative that avoids such limitations. Their practical success can be attributed to solid theoretical foundations based on VC-theory. SVM compute globally optimal solutions, unlike those obtained with ANNs, which tend to fall into local minima [4].

Least squares –support vector machines (LS-SVM) method was reformulated the traditional SVM algorithm. LS-SVM uses a least squares function with equality constraints which lead to a linear system which meets the Karush-Kuhn-Tucker (KKT) conditions for finding an optimal solution. Although LS-SVM simplifies the SVM procedure, the regularization parameter and the kernel parameters play an important role in the regression system. Therefore, it is necessary to establish a methodology for properly selecting the LS-SVM free parameters, in such a way that the regression obtained by LS-SVM must be robust against noisy conditions, and it does not need prior user knowledge about the influence of the free parameters values in the problem studied.

II. BACKGROUND

There are various tries to predict stock worth with Machine Learning. The main target of every research varies a great deal in 3 ways. (1) The targeting worth modification will be near-term (less than a minute), short-run (tomorrow to some days later), and semi-permanent (months later). (2) The set of stocks will be restricted to but ten explicit stocks, to stocks during an explicit business, to typically all stocks. (3) The predictors used will vary from international news and economy trend, to explicit characteristics of the corporate, to strictly statistic information of stock worth.

Based on the works we discover, a lot of progress has been created in predicting near-term and semi-permanent worth changes. These well operating models usually accept explicit info of the corporate that is commonly laborious to access by general public and doesn't work with short-run prediction.

We set to focus our project on the domain that presently has the worst prediction accuracy: short-run worth prediction on general stock victimization strictly statistic information of stock worth. Most researches during this domain have solely found models with around fifty to sixty % accuracy and that they usually

work just for classification. When studied thoroughly, we tend to noticed that some papers that claim to possess higher result used problematic metrics to guage the model, though the graphs or numbers may look fancy but the particular model provides restricted meaty steering for mercantilism.

III. LITERATURE SURVEY

Paper[1] :- The model trained by SVM shows pretty good effect with a high accuracy rate of 96.15%. A lot of parameters fit the training data perfectly well therby increasing the probability. The work design for the stock prediction system includes major factors that includes performance of the stock, dataset to predict the future performance, use Support Vector Machine (SVM) and classify it further to stock classification. Over-fitting in classifications and exaggerating minor fluctuations in the training data might decrease the subsequent predictive ability

Paper[2] :- This paper is based on the preliminary work that includes data preprocessing, data alignment and the use of Extreme Learning Machine algorithm which achieves good generalization performance at an extreme rate. The experiments are run with MATLAB R2010a in a PC machine which gives the prediction speed (running time) of each model at each time point (numbers in bold font indicate the best results at the given time point and the second best results are underlined). The bottleneck of any kernel based method is the definition of kernel 'k' that accurately reflects the similarity amongst all the data samples taken into consideration.

Paper[3] :- The proposed methodology can be summarized in the following modules – Data Collection, Feature Extraction Module , Training Module, Prediction Module . Support Vector Machine proved to be the most efficient and feasible model in predicting the stock price movement, in favour of the sentiments of the tweets. Through Machine Learning techniques for prediction purposes is inexpensive compared to the complete other survey that would have been conducted otherwise to gauge the public mood. Historical Twitter data cannot be obtained, unless it is saved by someone, so data has to be collected over a duration of a fixed number of months starting from the present date and time

Paper[4] :- In this research the stock market is tried to be predicted by three variants of Artificial Neural Network and by the Support Vector Machine algorithm which are Single Layer Perceptron, Multi-Layer Perceptron, Radial Basis Function, Support Vector Machines. It is evident from the result that SVM performed best on training set while the MLP algorithm did well on test data set. Conventionally, the prediction model works on test data set for the verification of the model that was built using the training data set.

IV. PROBLEM DEFINITION

Counting on one dataset might not be decent for the prediction and provide a result that is inaccurate conjointly having lower accuracy and taking longer for prediction. The previous projects existing provide lower accuracy with datasets not been taken from stock brokers.

Hence, we tend to take a collaborative approach towards this problem by integrating 2 hybrid models of Machine Learning- PSO-LSSVM (Particle Swamp Optimization and Least Square Support Vector Machine). Since PSO is straightforward to implement with larger accuracy and potency, once combined with LSSVM provides with shorter procedure time and can be ready to run on parallel computation

V. OBJECTIVE

The aims of this project square measure as follows: 1. to spot issue moving share market. 2. To come up with the pattern from giant set of knowledge of exchange for prediction. 3. To predict associate degree approximate worth of share value. 4. To produce analysis for users through net application.

The project are going to be helpful for investors to speculate available market supported the varied factors. The project target is to make net application that analyses previous stock knowledge of firms and implement these values in data processing rule to work out the worth that exact stock can have in close to future with appropriate accuracy. These foreseen and analyzed knowledge may be discovered by individual to grasp the money standing of firms and their comparisons. Company and business will use it to breakdown their limitation and enhance their stock values. It may be terribly helpful to even researchers, stock brokers, market manufacturers, government and general individuals.

VI. METHODOLOGY

AGILE methodology is a practice that promotes continuous iteration of development and testing throughout the software development lifecycle of the project. Both development and testing activities are concurrent. The following will be the system architecture:

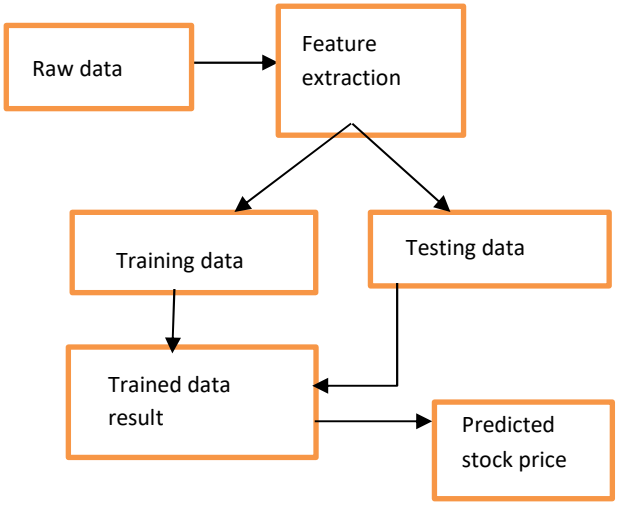


Figure No. 1: Basic Block Diagram

The training and test data will then be going through the hybrid module of PSO-LSSVM where the data goes through a technique of reaching more and more accuracy considering one head point. LSSVM algorithm is used for setting parameters for further processing.

LEAST SQUARE SUPPORT VECTOR MACHINE

Least squares support vector machines (LS-SVM) square measure statistical method versions of support vector machines (SVM), that square measure a group of connected supervised learning strategies that analyze knowledge and acknowledge patterns, and that square measure used for classification and multivariate analysis. During this version one finds the answer by finding a group of linear equations rather than a gibbose quadratic programming (QP) downside for classical SVMs. statistical method SVM classifiers, were planned by Suykens and Vandewalle..

Let X is n*p input data matrix and y is n*1 output vector. Given the {xi,Yi}i=1n,{ □ training data set, where xi E Rp andyi E R , the LS-SVM goal is to construct the function f(x)=y , which represents the dependence of the output i y on the input i x . This function is formulated as:

$$f(x)=Wt(x)+b$$

Where W and (x): Rp Rn are n*1 column vectors, and b E R. LS-SVM algorithm computes the function from a similar minimization problem found in the SVM method. But the main difference is that LS-SVM algorithm involves equality constraints instead of inequalities, and this algorithm is based on a least square cost function. In addition, the LS-SVM method solves a linear problem while conventional SVM solves a quadratic one.

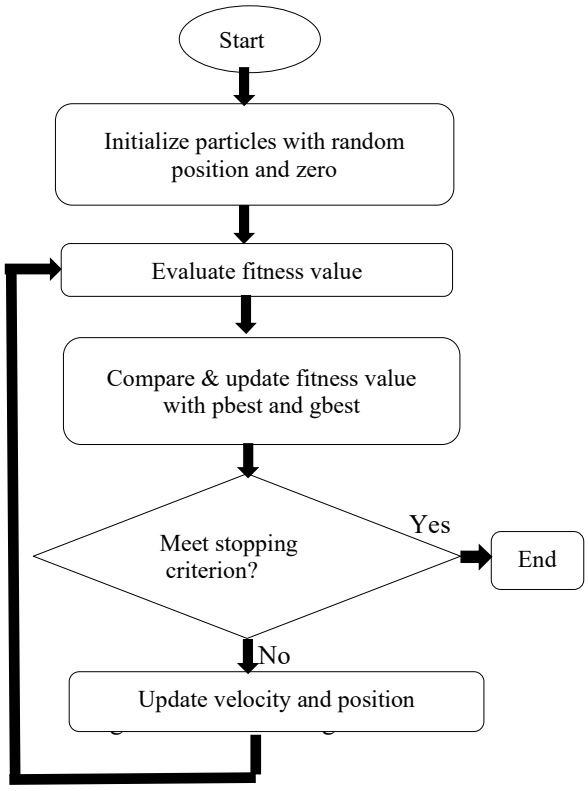
PARTICLE SWARM OPTIMIZATION ALGORITHM

PSO may be a comparatively recent heuristic search methodology that comes from the behavior of social teams like bird flocks or fish swarms. PSO moves from a group of points to a different set of points in an exceedingly single iteration with possible improvement employing a combination of settled and probabilistic rules. The PSO has been well-liked in academe and business, in the main as a result of its intuitiveness, simple implementation, and therefore the ability to effectively solve extremely nonlinear, mixed number optimization issues that area unit typical of complicated engineering systems. Though the "survival of the fittest" principle isn't utilized in PSO, it's typically thought-about as AN organic process algorithmic program. Optimization is achieved by giving every individual within the search house a memory for its previous successes, data concerning successes of a group and providing the simplest way to include this information into the movement of the individual .Therefore, every individual (called particle) is characterized by its position xi, its rate vi , its personal best position pi and its neighborhood best position pg. the weather of the speed vector for particle area unit updated as:

Vy□WVy+c1q(Xypb-Xy)+c2y(Xjsb-Xy),j=1,…,n

Where w is the inertia weight, xipb is the best variable vector encountered so far by particlei, and xsb is the swarm best vector, i.e. the best variable vector found by any particle in the swarm, so far 1 c and 2 c are constants, and q and r are random numbers in the range (0, 1). Once the velocities have been updated, the variable vector of particle i is modified according to:

xy□xy+vy,j=1,...,n.



VII. THE PROPOSED MODEL

The proposed system is based on the study of historical data, technical indicators and optimizing LS-SVM with PSO algorithm to be used for the prediction of daily stock prices. Levenberg-Marquardt (LM) algorithm is used as a standard benchmark for comparison with LS-SVM and LS-SVM-PSO algorithm. The proposed model contains six inputs vectors represent the historical data, derived technical indicators and one output represents next price.

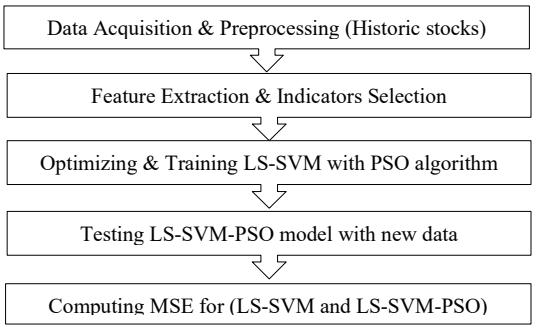


Figure No. 3: Proposed Model

The five technical indicators can be calculated from the raw datasets:

1. Relative Strength Index (RSI): A technical momentum indicator that compares the magnitude of recent gains to recent losses in an attempt to determine overbought and oversold conditions of an asset. The formula for computing the Relative Strength Index is given as follows.

$$RSI = 100- [100 / (1+RS)]$$

Where RS = Avg. of x days' up closes / Average of x days' down closes.

2. Money Flow Index (MFI): This one measures the strength of money in and out of a security. The formula for MFI is as follows:

Money Flow (MF) = Typical Price * Volume.

Money Ratio (MR) = (Positive MF / Negative MF).

MFI = 100 – (100/ (1+MR)).

3. Exponential Moving Average (EMA): This indicator returns the exponential moving average of a field over a given period of time. EMA formula is as follows.

$$EMA = [\alpha *T\ Close] + [1-\alpha* Y\ EMA].$$

Where T is Today's close & Y is Yesterday's close

4. Stochastic Oscillator (SO): The stochastic oscillator defined as a measure of the difference between the current closing price of a security and its lowest low price, relative to its highest high price for a given period of time. The formula for this computation is as follows:

%K = [(Close price – Lowest price) / (Highest Price – Lowest Price)] * 100

5. Moving Average Convergence/Divergence (MACD): This function calculates difference between a short and a long term moving average for a field. The formula for calculating MACD and its signal is given as follows.

MACD = [0.075*EMA of Closing prices] – 0.15]*EMA of closing prices]

Signal Line = 0.2*EMA of MACD

VIII. CONCLUSION

Machine learning is continually evolving with new strategies being developed each day. It's crucial that we tend to update our data perpetually and also the best thanks to do thus is by building models for fun comes like stock value prediction. Though the LSTM model higher than isn't adequate to be employed in live commerce, the foundations engineered by developing such a model will facilitate U.S.A. build higher models which may someday be employed in our commerce system.

Time series prediction could be a terribly intriguing field to figure with, as I actually have completed throughout my time writing these articles. There's a perception within the community that it's a posh field, and whereas there's a grain of truth in there, it's not thus tough once you get the suspend of the fundamental techniques.

Apart from Neural Networks, there is a unit several different machine learning models that may be used for commerce. The bogus Neural Network or the other Deep Learning model are simplest once you have over one hundred, 000 information points for coaching the model. This model was developed on daily costs to form you perceive the way to build the model. It's well to use the minute or tick information for coaching the model, which is able to offer you enough information for a good coaching.

IX. FUTURE SCOPE

The limitation of the proposed model is its computational speed, especially w.r.t. its sliding-Window validation as the computational cost increases with the number of forward day predictions. The proposed system does not predict well for sudden changes in the trend of stock data. This occurs due to many external factors and real-world changes affecting the stock market. We can although overcome this problem by implementing Sentiment Analysis and Neural Networks to enhance the proposed system. We can modify the same system to an online-learning system that adapts in real-time.

X. ACKNOWELEGMENT

We sincerely thank to our guide Mr. Sudhir Dhekane for his guidance and support for carrying out our project work and giving us new ideas to develop our project well. We also give our sincere thanks to our Principal Dr. B. K. Mishra, Vice-Principal Dr. Deven

Shah, HOD-IT Dr. Rajesh Bhansode, our project coordinators for arranging the necessary facilities to carry out the project work. We thank Thakur College of Engineering and Technology for their support and providing us a platform to explore new trending technologies in the world.

References

[1] Zhen Hu , Jie Zhu and Ken Tse ; Stocks Market Prediction Using Support Vector Machine ; 2013

[2] Feng Wang, Zhiyong Zhao, Xiaodong Li, Fei Yu and Hao Zhang ; Stock Volatility Prediction using Multi-Kernel Learning based Extreme Learning Machine;2014

[3] Tejas Mankar, Tushar Hotchandani, Manish Madhwani, Akshay Chidrawar,Lifna C.S ; Stock Market prediction based on Social Sentiments using Machine Learning ; 2017

[4] MehakUsmani, Syed HasanAdil, Kamran Raza and Syed SaadAzhar Ali ; Stock Market Prediction Using Machine Learning Techniques ; 2016

[5]Yunus Yetis; Halid Kaplan; Mo Jamshidi (2014), "Stock Market Prediction by Using Artificial Neural Network", 2014 World Automation Congress (WAC), pages: 718 - 722, DOI: 10.1109/ WAC. 2014.6936118.

[6]MehakUsmani; Syed HasanAdil; Kamran Raza; Syed SaadAzhar Ali "Stock Market Prediction Using Machine Learning Techniques" 2016 3rd International Conference on Computer and Infor- mation Sciences (ICCOINS) Pages: 322 - 327, DOI: 10.1109/ICCOINS.2016.7783235

[7]Adebiyi Ayodele A., Ayo Charles K., Adebiyi Marion O., and Otokiti Sunday O. (2012), "Stock Price Prediction using Neural Net- work with Hybridized Market Indicators", Journal of Emerging Trends in Computing and Information Sciences, Volume: 03, Issue: 01, Pages: 1-9.

[8]RaduIacomin (2015), "Stock Market Prediction", 2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, CheileGradistei, Romania, Pages: 200 - 205, DOI: 10.1109/ICSTCC.2015.7321293.

Supervising heart disease using machine learning Algorithm

Amish Punmiya, Ketan Samrit, Saish Shetty, Shridhar Kamble
Thakur College of Engineering and Technology, Mumbai, India

Abstract - **Recent studies have shown that the world is encountering a high rate of heart ailments. Heart diseases have become one of the leading causes of death, and World Health Organization (WHO) states that cardiovascular diseases are the world's largest killers, which claimed 17.1 million lives a year. India having a population of approximately 1.33 billion people, more than one sixth of the world's population, experiencing heart disease as the single largest cause of death in the country with heart attacks being responsible for one-third of all death caused by heart diseases. Monitoring person's behaviour at some stage in sleep, specially heart rate and breathing fee has obtained greater attention from the last few years. Many systems have been advanced to display, discover chance of coronary heart failure below sure situations. Most of these structures screen important signs and symptoms with the aid of measuring one or greater components of the ballistic pressure during a heartbeat pulse, ranging from force magnitude, stress, to the ensuing function trade. Even though they may be able to carry out correct tracking, most of them are pretty bulky to put in or are inconvenient to the customers. Advanced technology like this system makes it possible for the doctors to get information about the patient's heart beat using it without interrupting their life. Continuous heart monitoring is useful for capturing irregular heart beat that happen infrequently. This may then be useful for performing further actions which are necessary to keep the heart beat at regular level.**

Keywords - IOT, ML, Health Monitoring, Arduino, Node MCU, Heart disease prediction, Heart rate sensors, Health diagnosis.

I. INTRODUCTION

The Internet of Things (IoT) is a communication of embedded devices using networking technologies. IoT will be one of the important trends in the upcoming years. It will affect the networking, communication and business. By keeping all these things in minds, technology really proves to be helpful for an individual. With new technological advancement, lots of smart sensors or medical sensors have been introduced that continuously analyse individual user activity and automatically predict a heart attack before the individual feels sick. Therefore, identifying the appropriate sensors is important. Nowadays patient is more concerned and takes actively part in collecting and reviewing his reports.

In today's world, various wireless communication standards have enabled the sensor to develop from traditional forms i.e. require active patient involvement to passive form i.e. require no involvement from user. Elderly people got to make regular visit to the doctor for his or her health signs test results. Observing on regular basis of essential signs is compulsory as they're main signs of well-being of one's individual health. These vital signs include,

a. Pulse rate

b. Body temperature [1]

The goal is to develop a low power, more reliable technology that analyses essential signs, and continually monitors users by gathering information from the body and send the parameters through wireless technology. There must be a specified method of transmission and to display the signal after the data is recorded.

There are many technologies which help us to monitor the patients who are not in reach of common clinical settings, which will increase the access to care and reduce the cost of delivering human healthcare. In many developing countries, Human Healthcare is usually defined as a serious problem i.e. almost as much as 8% of the population who can has access to 20% of the country's medical resources. This inability to access proper Human Healthcare for rural populations paired with rapidly increasing disorder rates poses a significant problem. This paper proposes a remote monitoring and sensing parameter of the human body which consists of pulse and temperature. The parameters that are used for sensing and observing will send the data through wireless network using firebase as database. Adding a web-based observing helps to keep track of the regular health status of a patient. The sensing data will be collected in database continuously and will be used to inform the doctor and the patient about the current trend in health. [1]

II. BACKGROUND

There are about 55 million annual deaths due to cardiovascular diseases in India every year. These are mainly due to the lack of knowledge among the masses. The Indian people tend to take things too lightly in context of diet and maintaining fitness. Hence there is an above average rate of people falling prey to heart diseases. Nowadays better technologies and better guidance from physicians is resulting in a control on the resulting deaths from cardiovascular diseases.

Many heart rate sensing devices are available in the market, and they give high accuracy rate in providing the pulse rate. Fitbit Versa is a wrist band watch used for sensing the heart rate and pulse rate of the body and it also tracks other activities, for instance while jogging it counts the kilometres and the footsteps of the user. Secondly, Fitbit band also plan the user daily activity which has to be performed. This activity is based on the past records. Apple Watch Series 1 is a kind of the wrist watch which sense the heart rate, maintain some diagnostics and integration to this

it is also used for picking up the call and showing content of the iPhone, iPad, and iMac on the watch. These devices are used only for sensing the heart rate while the device is on the hand but our system which can sense the ballistic force with the help of the sensors.

III. Objectives and scope of project

The project aims at creating a system which is able to sense the heart rate using IR sensor and Pulse Sensor and the pulse band sensor intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other user suffering from heart diseases. This also aims at having a basic machine learning capability for heart attack prediction. In an era where medical diagnosis and treatment has made massive leaps, it is unfair to deny anyone proper healthcare owing to a lack of means.

One often sees a lack of a proper software to help it out in such cases. Software can avoid the necessity of having to revisit doctors, or to spend hours self-researching complex medical terms and appendices, as many diagnosed patients tend to do. The use of a band here can calculate the rate every minute and pass the data to the database. [2] This system can be used for providing enhanced healthcare services to the rural areas of the developing countries which are facing shortage of efficient specialized doctors. Thus, the cardiovascular disease causing the death of patient can be reduced immensely by the implementation of this proposed system, in both rural and urban areas.

1) Heart hate monitoring will help patient understand the nature of the body system.
2) Cost Effective system with multiple features embedded on a single unique system.
3) Wristbands and geophones help in early detection of the heart attack or stroke and help save lives.
4) It helps in alerting you and your near ones about sudden changes and help can be provided in medical emergency.

Besides the features above, we also aim to make the software robust, accurate and fast.

IV. Working of the project

The proposed system consists of sensors which are used to sense bed vibrations caused by the ballistic force. The sensors help in monitoring person's heart rate during sleep. This mechanism for heart rate monitoring does not require a special kind of bed. These sensors are placed at the side of the bed, which allows person to sleep in a comfort manner as well as person can move around and change position freely during sleep. The sensors are integrated with Arduino, which connected to the ac amplifier. The vibrations sense by the vibration sensors are converted from digital signal to analog signal which is to be displayed on to the monitoring screen.

This system can be used for providing enhanced healthcare services to the rural areas of the developing countries which are facing shortage of efficient specialized doctors. Thus, the cardiovascular disease causing the death of patient can be reduced immensely by the implementation of this proposed system, in both rural and urban areas. Heart hate monitoring will help patient understand the nature of the body system. Cost Effective system with multiple features embedded on a single unique system. [7] Wristbands and vibration sensor help in early detection of the heart attack or stroke and help save lives. It helps in alerting you and your near ones about sudden changes and help can be provided in medical emergency. Besides the features above, we also aim to make the software robust, accurate and fast.

The project basically consists of two modules. The modules are developed separately to be kept at the doctor's side and patient's side. The first module would be placed at the doctor's side, when a patient approaches the doctor when he feels unwell, the doctor would need to examine the possible reason for illness. The proposed system would prove of major help for the doctor in doing so.

The module 1 of the system would consist of a pulse sensor and a temperature sensor that would be mounted on an Arduino board and would send the data to firebase database. The data from firebase would be retrieved by a healthcare management system which the doctor can access. The doctor can add the patient data for a specific patient in the online system. The module 2 would be a patient side system that would consist of a wearable pulse sensor mounted on a node MCU. The node MCU would be connected to the internet and used to send the pulse data to the firebase database, the data from firebase would be displayed on an android application. [4]

This android application would also have an option to set the lower and higher acceptable value of the heartbeat. If the heartbeat gets out of the specified range the IFTTT URL would be triggered. This would send a notification and an email to the concerned people like the doctor and the closest relatives of the patient. The second part in module 2 would be deployment of vibration sensors to the bed of the patient having chances of heart attack. It is based on the study that generally before a probable heart attack, the body generates vibrations at a specific frequency. These vibrations would be detected by the vibration sensors and a buzzer would be sounded to alarm the patient so that preventive steps could be taken. [3]

V. Summary

Healthcare, though a new venture for the Artificial intelligence/ Machine learning domain, remains one of the most crucial domains of public service. With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care, and community services. It is pertinent that we utilize mathematical modelling to iron out any imperfections and aberrations, so that we have the rigorous standards of accuracy that the medical diagnosis field requires.

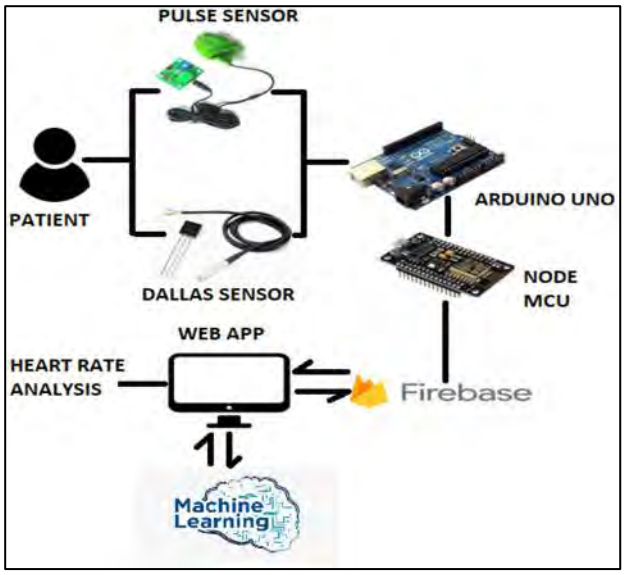


Figure 6.1 Heart Monitoring & Analysis

Many virtual assistants are available on the market, and they give high accuracy in handling input, particularly Siri and Google Assistant. Although not anywhere close in accuracy, the application understands most commands with enough accuracy to extract results through them.

The project aims at creating an intelligent but lightweight healthcare assistant. In an era where medical diagnosis and treatment has made massive leaps, it is unfair to deny anyone proper healthcare owing to a lack of means. This project aims to make healthcare less exclusive than it currently is and make it easier to find and avail services. This application will be able to help healthcare and other hospital staff work efficiently.

Stakeholder perspective is crucial to the success of this project. It is also vital to involve the patients, as the patients being comfortable with the equipment, both hardware and software, is necessitous to the smooth working of the technology. Potential users seemed open to the idea, as anticipated. Stakeholders here include - Doctors and nurses, Health insurance companies, Hospitals, Health department (State as well as National/Federal), Biotechnology manufacturers and even patients.

This application will be able to help healthcare and other hospital staff work efficiently. This also aims at having a basic machine learning capability for disease prediction.

The main objective is to offer a way for doctors and patients alike to easily perform mundane healthcare activities as well as advanced prediction services.

VI. Design and Implementation

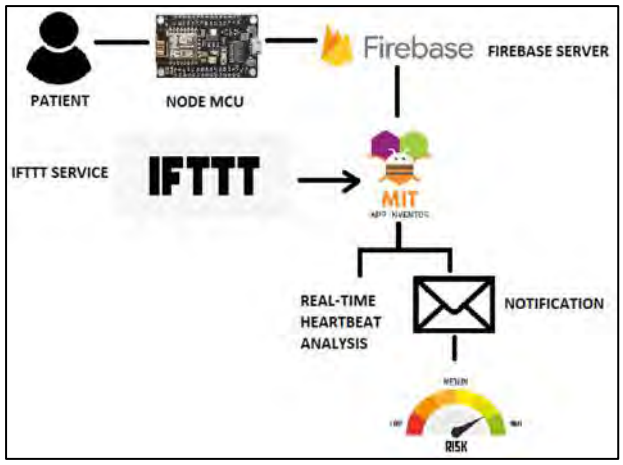


Figure No 6.2 Heart Rate Detection and Monitoring using Pulse Sensor Band.

VII. Deployment

Deployment is a major challenge in software and hardware which involves machine learning and software which will help in detecting the early heart attack from the given pulse rate and heart rate which is collected by vibration sensor and wrist pulse band. Huge amounts of processing are required just to get it up and running. Data generated in the process is humongous as well. The first module of the project would be deployed on the doctor's side. It would not be much of a task. The second module would be deployed individually at the customers end and hence would require some technical work like setting up the firebase database for every customer etc.

The GUI by itself is not very difficult to deploy. We have multiple ways of offering interfaces, such as web interfaces and command line interfaces. It will expand to

Figure No 6.3 Pulse sensor mounted on Node MCU

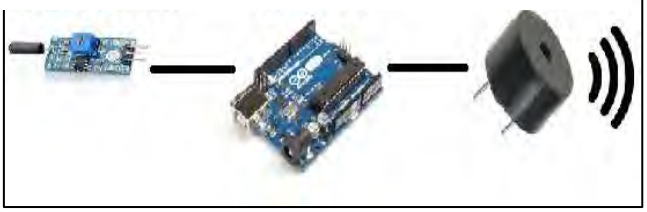


Figure No 6.4 Vibration Sensor Based Heart Attack Detection.

a desktop GUI in the future as well. The Web interface will be deployed using Somee.com, depending on the relative cost of hosting.

The command line can be used as a windows executable file. Users can simply download it and run it natively on their windows PCs.

Maintenance of the system is also a major challenge for the developer. As the system would continuously gather and store data the storage would be a major concern in case of online database.

VIII. FLOWCHART

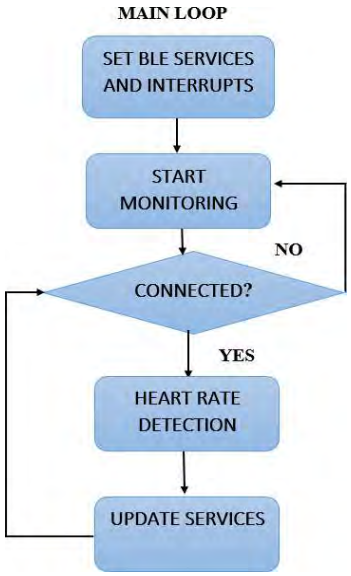


Figure No 8.1 Flow Chart for Heart rate detection.

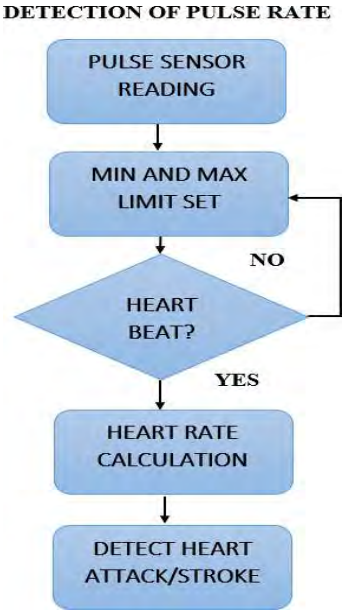


Figure No. 8.2 Detection of pulse rate

IX. RESULTS AND DISCUSSION

A. Output

We were able to build a pulse sensing wristband which is able to sense the pulse rate of the body. Besides this, we were also able to add some basic features such as transferring that data to the Node MCU and uploading the data to firebase server. The data from firebase would be displayed on the webpage hosted on the internet. The data from firebase can also be displayed on an android application for the user's convenience.

B. Outcomes

The project correctly analyses the pulse rate using pulse band sensor and displaying it on the screen. It uses Node

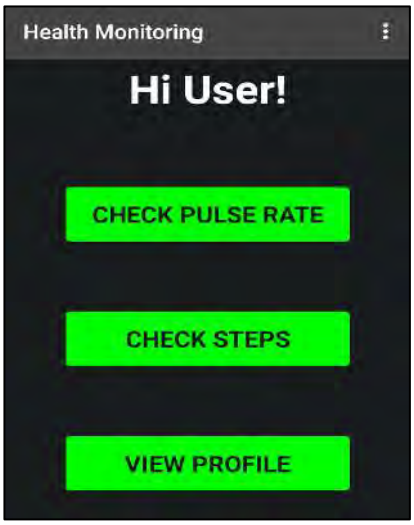


Figure No 9.1 Application Interface

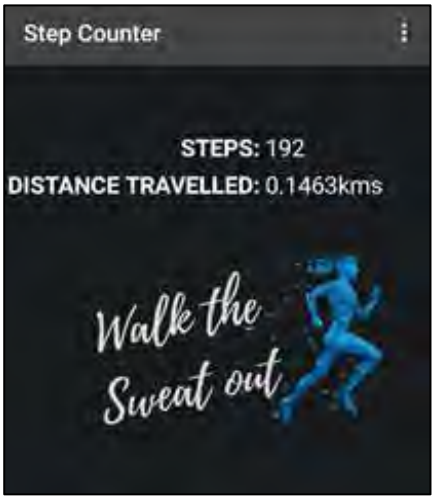


Figure No 9.2 Step counter

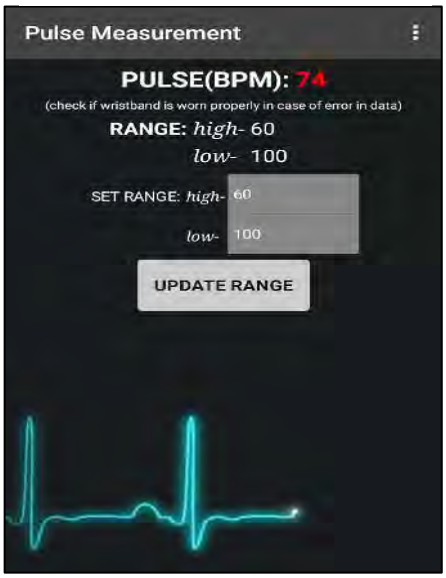


Figure No 9.3 Pulse Measurement

MCU which is for transferring data from the sensor to the

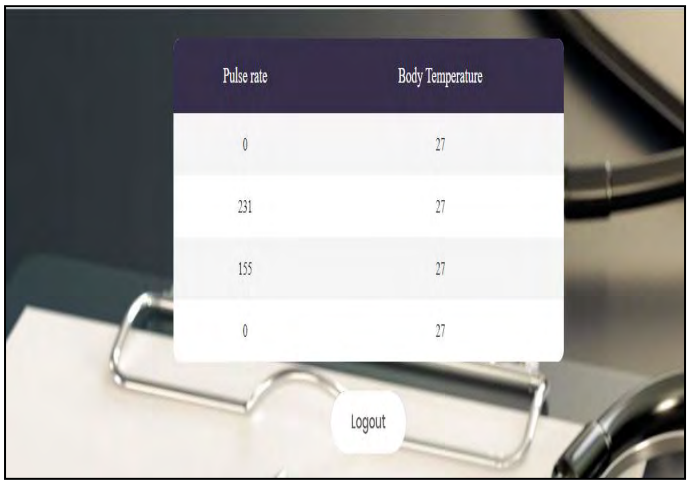


Figure No 9.5 Firebase data

device. The data is stored in firebase database. Data is retrieved from firebase and displayed on the doctor's dashboard. This data can be used for further analysis by the doctor in predicting the disease of the patient. The dataset that would be available would be tested using pretrained data and would be used for drawing important conclusion useful for the doctor. The continuous monitoring of the patient and availability of historic data would enable the doctor to make a better study of the patient's health condition.

We were able to build a pulse sensing wristband which is able to sense the pulse rate of the body. Besides this, we were also able to add some basic features such as transferring that data to the Node MCU and uploading the data to firebase server. The data from firebase would be displayed on the webpage hosted on the internet. The data from firebase can also be displayed on an android application for the user's convenience.

C. Discussion

AI systems are now a part of a variety of specializations from stock trading to the setting of real estate prices; from fraud detection to translation of vernacular languages; from designing our daily shopping lists to forecasting which movies we might like and diagnosis of medical condition. AI is gradually replacing the work that would require about hundreds of workers for its completion and hence is proving a major threat to jobs in future.

This is merely the beginning. Not long from today, AI will be used to advance our understanding of biological phenomenon through analysis of highly dense and huge datasets, help us invent new drugs and make treatments more thorough and personal. Self-driving vehicles can revolutionize transportation and allow new paradigms in planning of architecture. Machines will automate our homes more efficiently, make businesses more productive and help predict risks to society.

Before Artificial Intelligence systems can be deployed in healthcare applications, they need to be 'trained' through datasets that are generated from clinical activities, such as screening, diagnosis, treatment assignment and so on, so that they can learn similar groups of subjects, link between subject features and outcomes of interest. Use of AI in future would gradually lead to increase in accuracy and efficiency of the system.

The main focus in an AI application should be its accuracy in predicting the result. If the system has a good idea but has poor accuracy, it is of no use. The system especially if being used in healthcare field, must have an accuracy of over 95 percent. The accuracy of the system can be increased by training the system using large datasets.

X. CONCLUSION

Our system helps to detect the heartbeat of the patient who will be sporting/wearing the sensors. Whenever there is an irregular heart rate detect by the sensor it will be alerted to the patient, doctor and their relative via the IFTTT notification. Through which further actions can be taken to restore the normal heart beat of the patient. We developed a system that measures and detect heartbeat and body temperature of the patient, sends the info to user or server end by using microcontroller with reasonable cost and great effect. Use two different sensors (pulse sensor and Dallas temperature sensor) and these are mainly under the control of microcontroller. For Human Heartbeat measurement use fingertip, it's in bpm (beats per minute). These calculated rates will have stored in server by transferring through Wi-Fi module via internet. To measure the human body temperature, used a Dallas temperature sensor, the measured data is given to transmitter module, it in turn transfer these data to firebase database through wireless system due to this notice avoided use of wires.

Finally, the stored data in server will be displayed for further analysis by physician or specialist to provide better

aid. The data would be continuous and would be stored at 30 second time intervals. The data would also be available in its historic form and would provide the physician a better idea of the patient's overall condition. All the data would be available at the doctor's dashboard that can be accessed from anywhere and anytime. Also, machine learning is applied on the dataset for automation of the process.

XI. Future Scope

Currently the project focuses on gathering a structured authentic data. As medical data and analysis is a sensitive field, the project intends to find an assimilation of the results of various algorithms and compare the accuracy. It's a challenging task for monitoring and detect heart rate and predict heart attack/ stroke from pulse rate itself. There are other parameters also by which heart attack/ stroke may occur such as patients diabetic, blood pressure history and all. So, we can consider these parameters also to develop enhanced health monitoring system.

The project aims at creating a system which is able to sense the heart rate using IR sensor and the pulse band sensor which is intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other user suffering from heart diseases.

Our system helps to detect the heartbeat of the patient who will be sporting/wearing the sensors. Whenever there is an irregular heart rate detect by the sensor it will be alerted to the patient, doctor and their relative via the IFTTT notification. Through which further actions can be taken to restore the normal heart beat of the patient.

Proposed health monitoring system will monitor the heart beat fluctuation using sensors and monitoring, detection and analysis will be done using pulse-wrist-band which is equipped with Node MCU, Pulse Sensor**.**

This system will help the patient to detect the irregular heart beat within stipulated time. The real time monitoring will avoid any fatality depending on the further actions taken.

XII. Acknowledgment

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have completed the project successfully. I would like to thank everyone for their guidance.

We sincerely thank our principal, Dr. B.K. Mishra and HOD, Dr. Rajesh Bansode for always encouraging us to do our best. We highly indebted to our guide Mrs. Shridhar Kamble, who supported and constantly supervised us throughout the project and helped us in not only completing this project but also provided us with ample amount of knowledge that was really beneficial to us.

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all teaching staffs of IT Department who helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in laboratory for their timely support.

References

[1] Gowrishankar S., Prachita M. Y., Arvind Prakash, "IoT based Heart Attack Detection, Heart Rate and Temperature Monitor", International Journal of Computer Applications (0975 – 8887) Volume 170 – No.5, July 2017

[2] Rajalakhshmi.S S. Nikilla," Real Time Human-Health Monitoring System using Arduino", South Asian Journal of Engineering and Technology Vol.2, No.18 (2016) 52– 60.

[3] Z Jia, A Bonde, S Li, C Xu, J Wang, Y Zhang, RE Howard, P Zhang, "Monitoring a Person's Heart Rate and Respiratory Rate on a Shared Bed Using Geophones", Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems, 2017.

[4] Aida Kamišali´c, Iztok Fister, Jr, Muhamed Turkanovi´c and Sašo Karakati˘c "Sensors and Functionalities of Non-Invasive Wrist-Wearable Devices: A Review", Sensors 2018, 18, 1714; doi:10.3390/s18061714, 25 May 2018

[5] Chao Li, Xiangpei Hu, Lili Zhang, "The IoT-based heart disease monitoring system for pervasive healthcare service", ELSEVIER Procedia Computer Science, Volume 112 Issue C, Pages 2328-2334, September 2017.

[6] Zhenhua Jia, Richard E. Howard, Yanyong Zhang, Pei Zhang, "HB-Phone: A Bed-Mounted Geophone-Based Heartbeat Monitoring System", Proceeding of 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Pages 275-276, 2016.

[7] Nicholas Constant,Tanya Wang, Kunal Mankodiya, "Pulseband: A hands-on tutorial on how to design a smart wristband to monitor heart-rate", IEEE Virtual Conference on Applications of Commercial Sensors (VCACS), April 2016.

[8] Prosanta Gope and Tzonelih Hwang, "BSN-Care: A Secure IoT-Based Modern Healthcare System Using Body Sensor Network", IEEE SENSORS JOURNAL, VOL. 16, NO. 5, MARCH 1, 2016.

[9] Grace Wusk, Hampton Gabler "Non-Invasive Detection of Respiration and Heart Rate with a Vehicle Seat Sensor", doi: 10.3390/s18051463, May 2018.

[10] Online Web Resource referred on 5-08-2018
https://www.livemint.com/Science/6j9LCCGcUiOVgG0gHCkr0O/25-million-died-due-to-pollution-in-India-in-2015-Lancet-s.html

Student Community Portal

Siddhesh Naik, Jay Mithani, Kartik Adak, Vinay Jain, Vandana Munde
Information Technology Thakur College of Engineering and Technology Mumbai,India
siddheshnaik2412@gmail.com, mithani.jay98@gmail.com, kartikadak98@gmail.com, hncvj@engineer.com, vrsakharkar@gmail.com

Abstract - **Student discussion portal are considered an extension of traditional learning that promotes dialogue, reflection, knowledge construction and self-assessment. This project represents the design and development of student discussion portal in which the community is limited to students studying in particular college registered students can raise discussions and topics and students can reply to that discussion with remark, comment or a rating and based on that rating the trending discussions will be displayed in the timeline. Now students can reply to particular registered user privately. Addition to that students can upload documents related to particular topic and any registered intended user can download it. Students can login whenever they want and upload/raise a discussion. Users can search according to tags on discussions.**

Keywords – extension of traditional learning, Timeline, tags

I. INTRODUCTION

Discussion forum also commonly referred in many kinds of the terms like web forums, message boards, discussion boards and discussion groups. Members can organize the discussions and gather the suggestions through online around various people. The discussion is initiated through any one member by posting a topic and other members of the group cam make reply. The authorized members of the same group can able to share information, suggestions and ideas among the group. The features in online course were that gives the instructor the opportunity to post topic for discussions and providing the students a tool for responding each topic is referred to as "thread". Technology wherever users will move with different user by exchanging messages generally referred to as post, and discussing a given subject (topic). A connected cluster of written conversations in a web setting. A forum might even be a neighborhood, situations, or group in which people exchange ideas and discuss issues, especially important public issues. A discussion board forum contributes to the event of your psychological feature and necessary thinking skills, permits times for thoughtful, thorough reflection on the right track topics, facilitates searching learning by allowing you to review and respond to the work of others and paves the way for you to approach own learning.

An online discussion forum is the based communication tool (or activity) that enables is participants to post messages and to reply to others messages asynchronously. Over time discussion become increasingly user friendly. Forum usually organized messages by thread topic, date, and time and allowed users to retort to determine message or a produce be a replacement message, of their own. in order to become part of a forum, many sites required web as to register for a user id and password. College professors began utilizing these forums as teaching tools. Students were encouraged to use specific forums set up by the college or professor to discuss the topics. The largest cluster of dialogue forum users, however, was created from people seeking information. Online discussion forum (ODF) is a web-based application that has been used extensively to bring people together with shared interest and mind-set. The use of online discussion forum (ODF) has emerged as a common tool and an effective way of engaging students outside the classroom [1]. ODF is an e-learning platform that permits students to post messages to the discussion threads, interact and receive feedback from other students and instructor, and foster deeper understanding towards the subject under study. In education, they have been deployed to complement traditional techniques such as lectures and tutorials [2]. Online discussion forums (ODFs) harmonize with the educational philosophy that considers communication a necessary and fundamental mechanism for effective learning. It was discovered that learners' interaction with both human and inanimate objects, and their participation in technology mediated education, were essential for the standard of their learning experience, which can enrich the process of knowledge exchange among participants and has positive effects on the students' performance [10]. Consequently, online discussion forums (ODFs) can be successful in enhancing collaborative learning by attracting students to participate and interact [8].

II. PROBLEM DEFINITION

With extremely growing of the telecommunication infrastructures, like net and also the development of the

high performance has led to new era of fast advances in data technology. Internet has become associate ocean of knowledge associated with each side that has existence during this world. The main purpose of this web site is to develop a 1 roof platform for the effective interaction, effective exposure, and a right direction toward communication. Main aim is to provide our students an opportunity to enhance their knowledge by sharing their views on this platform by having discussions with other users. This web portal will act as a healthy questionnaire which helps to give and get effective solutions with best of their experience.

III. BACKGROUND

The online student discussion is one of the key assessment components and requires students to answer the questions. The objective was to help students demonstrate their ability to work collaboratively using online discussion forum moderated by the lecturer and resolve their differences in developing an optimal solution collecting via synergies generated from collective wisdom and peer-to-peer learning. It requires students to participate in an online discussion forum on a weekly basis in a semester. During this time, they are expected to demonstrate their individual ability to apply the concepts and theories of enterprise systems to a business organization and suggest workable solutions to the problems/issues raised in a comprehensive case study. A tutorial may be a period of sophistication outside the 'regular lecture classes' where lecture material is discussed between students and an educator assistant or a professor. Communication in the electronic tutorial was asynchronous. No between-group differences were found regarding grades. From his assessment, students seemed to prefer the electronic tutorial due to the flexibility it offered and the fact that a permanent record was available for future reference. A study of online forum usage during a graduate science education course showed that contributions by students, were more detailed and deliberate when interacting online compared to face-to-face interactions [7]. Online forums provide an avenue where class material can be reviewed, discussed and reinforced. The underlying presumption is that knowledge is often constructed by a web asynchronous dialogue of sophistication material. The conversational model of learning – Laurillard [9] – stipulates that it'll enhance learning, including increased motivation, engagement, and deeper levels of understanding. Communications technologies – like online forums – enable students to debate class material in an asynchronous manner. This channel supports discussion outside the classroom and is flexible for college kids that employment full-time. [5] used a constructivist interaction analysis model to evaluate learning in an online community. However, their analysis revealed that most of the online interactions were limited to the social interchange rather than the creation of knowledge. The authors indicate that this finding was a product of their choice of forum – an information-share forum was used rather than an academic forum.

IV. SIGNIFICANCE

Personal responses to forum topics aren't restricted in time or the length. Students have the freedom to continue dialogues about the topics that interested them most. Effective forum topics are open-ended and designed to encourage students to require a grip on problems. To respond to a forum topic needs organized thought and synthesis of ideas introduced in school. If a view of student is challenged, he or she typically adds carefully considered reasons to back up previous comments. Students have the flexibleness to mirror on their thoughts and skim the responses of others. Many students have declared that they habitually mentioned the forum topics with friends, family and colleagues outside of sophistication before golf stroke their comments on-line. Students have long been want to class discussions within the confines of a brick-and-mortar classroom. albeit a proper discussion doesn't occur, there's always a touch of informal time with their teacher and classmates to debate and gauge their understanding of the category material. Understandably, discussion forums are an integral a part of a web class; however, students routinely voice their annoyance with- or downright hatred of- these forums. While some students would rather add isolation and just get their work done, most students would really like to urge more out of their class. Why do they resent discussion forums? Could it's that they sense the shortage of authentic and meaningful interaction? Could it's that they view it as just another assignment with a "right" and "wrong" answer? In some instances, they might be spot on. While it's important to possess forums for discussion, it's even as important to line them up during a meaningful way. A successful class discussion format creates a way of community, enabling students to attach with each other. This connection not only helps students better enjoy the web experience; it also leads to a deeper investment within the class. This intrinsic motivation is vital to overall success. Additionally, a carefully-constructed student discussion allows the teacher to select abreast of misunderstandings or weaknesses

which will be proactively addressed with the whole class.

V. PROPOSED SYSTEM

In student community portal a student is fundamentally an understudy who will utilize the dialog entrance for setting up any inquiries or questions he/she may experience. On the off chance that an understudy is new to the portal, at that point he/she needs to enroll first before making any talk they need or else they are just permitted to see the exchange of different understudies however can't reply to that dialog except if registered. Once getting themselves registered, they can just login to the community portal from at whatever point next time they need to start the discussion. Client credential (Id and secret phrase) are along these lines put away into the database while enlisting and cross-checked from a similar database while signing in. Aside from this, administrator can affirm/erase the dialog dependent on content against network rules being utilized.

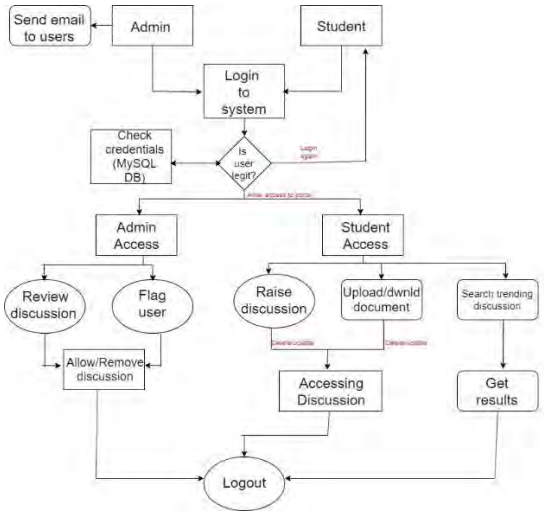


Figure 1. Work-Flow

Data Flow

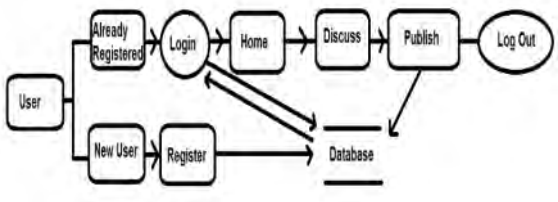


Figure 2. Data Flow

A User of the portal is basically a student who will be using the discussion portal for putting up any questions or doubts he/she might encounter. If a student is new to the system then he/she needs to register first before creating any discussion they want or else they are only allowed to view the discussion of other students but cannot answer to that discussion unless registered. Once getting themselves registered, they can simply login to the system from whenever next time they want to access the portal. User credentials (Id and password) are subsequently stored into the database while registering and cross-checked from the same database while logging in. Apart from this, admin can approve/delete the discussion based on content against community guidelines being used.

Some of the common features of forums are discussed below:

- Private message: a personal message, or PM for brief , may be a message sent privately from a member to at least one or more other members. the power to send so-called carbon copies is sometimes available. When sending a carbon (cc), the users to whom the message is sent directly won't remember of the recipients of the carbon or maybe if one was sent in the first place.
- Attachments: An attachment are often almost any file. When someone attaches a file to a post they're uploading the file to the forums' server. Forums usually have very strict limit on what are often attached and what cannot (among which the dimensions of the files in question). Attachments are often a part of a thread, group, etc.
- Emoticons: An emoticon or smiley may be a symbol or combination of symbols want to convey emotional content in written or message form. Forums implement a system through which a number of the text representations of an emoticons (e.g. xD, :p) are rendered as a little image. counting on part the planet the forum's topic originates (since most forums are international) smilies are often replaced by other sorts of similar graphics, an example would be kaonic(e.g. *(^O^)*, (^-^)b), or maybe text between special symbols (e.g. :blink:, :idea:).
- Poll: Most forums implement a poll system for threads. Most implementations leave single-choice or multi-choice (sometimes limited to a particular number) when selecting options also as private or public display of voters. Polls are often Electronic copy. Members choose a poll and a statistic is displayed graphically.

VI. OBJECTIVES

The aims of this project are as follows: Developing a closed student community web portal which is limited to specific student group or particular college. To achieve a flexible audacious student level network. Enable students to raise currently trending topic related to activities happening in college. Online discussion forum (ODF) could be a web-based application that brings individuals along with shared interest and mind-

set. The use of online discussion forum (ODF) way of engaging students outside the classroom1. ODF is associate degree e-learning platform that has students with privilege to post messages to the discussion threads, interact and receive feedback from other students. Online forums offer associate degree avenue wherever category material is reviewed, discussed and reinforced. The underlying presumption is that knowledge can be constructed by an online asynchronous dialogue. The main feature of the project is a web portal where students can raise discussions, privately message other users, upload files and documents limited to college and expect a secured layer gateway to enter the portal. An online discussion forum is the based communication tool (or activity) that enables is participants to post messages and to reply to others messages asynchronously. While we tend to consider discussions as face-to-face activities, online discussions provide benefits sometimes not available in the classroom.

For example:

- More students are able to participate in online forums than in large group classroom discussions.
- It provides the platform for students who do not participate during class discussions to become “vocal” during online discussion.
- Online interaction allows for further discussion of a topic. It also provides students with enough time to think and reflect on the course material and hence post more thought-out comments to the discussion board.
- Students and instructors often think of comments or questions after the lecture and can post them and discuss these inquiries online.
- The online environment lets you easily create small discussion groups within a large class.
- The online format is a convenient way to archive and preserve your course discussions.
- Online discussions can give you feedback prior to or after a class; this feedback may indicate what content students understand and what requires further clarification.

VII. METHODOLGY

AGILE methodology could be a apply that promotes continuous iteration of development and testing throughout the software system development lifecycle of the project. Both development and testing activities are concurrent. The agile software system development emphasizes on four core values. Individual and team interactions over processes and tools two. Working software over comprehensive documentation. Customer collaboration over contract negotiation. Responding to change over following a plan. Agile development with Scrum is the light at the end of the tunnel.

Scrum is a subset of Agile. It is a lightweight procedure structure for light-footed improvement, and the most generally utilized one. A "process system" is a specific arrangement of practices that must be followed all together for a procedure to be predictable with the structure. (For instance, the Scrum procedure system requires the utilization of improvement cycles called Dashes, the XP structure requires pair programming, etc.) A Scrum procedure is recognized from other coordinated forms by explicit ideas and practices, isolated into the three classes of Jobs, Ancient rarities, and Time Boxes. These and different terms utilized in Scrum are characterized beneath. Scrum is regularly used to oversee complex programming and item advancement, utilizing iterative and gradual practices. Scrum essentially expands profitability and diminishes time to benefits comparative with exemplary "cascade" forms. Scrum forms empower associations to alter easily to quickly evolving prerequisites, and produce an item that meets advancing business objectives.

VIII. RESULTS ACHIEVED

Students are now able to understand WordPress Framework and its fundamentals. This will help them to understand more of the development side of website and technologies being used to connect to database and retrieve from the same to check credentials of registered students. Apart from this, students also enable themselves to understand fundamentals of a discussion forum where they can raise a discussion based on the query they have and other student who had already registered can view the discussion and answer to discussion of other students. This ultimately facilitate them to urge their queries resolved, get discovered by different students, share their thoughts and additionally ping one another (if each have accepted friend request of every other) and additionally transfer study materials in style of document. More students are able to participate in online forums than in large group classroom discussions. It provides the platform for students who do not participate during class discussions to become “vocal” during online discussion. Online interaction permits for additional discussion of a subject. It also

provides students with enough time to think and reflect on the course material and hence post more thought-out comments to the discussion board. Students often think of comments or questions after the lecture and can post them and discuss these inquiries online. The online environment lets you easily create small discussion groups within a large class. The online format may be a convenient thanks to archive and preserve your course discussions. Online discussions can give you feedback prior to or after a class; this feedback may indicate what content students understand and what requires further clarification.

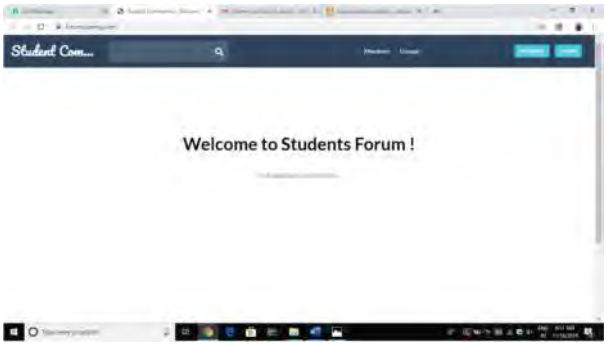


Figure 1. Home Page

Figure 2. Registration Page

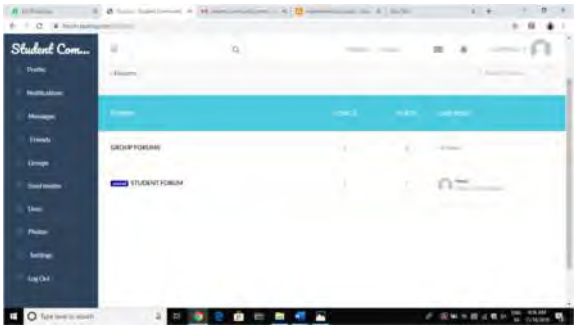


Figure 3. Discussion Page

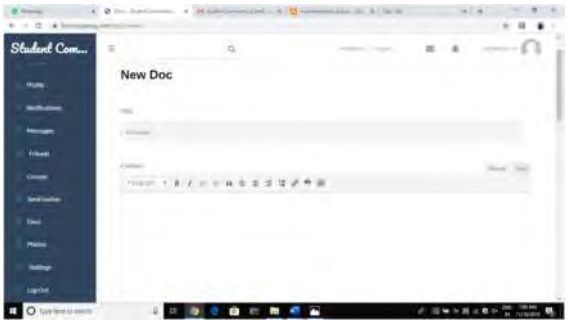


Figure 4. Document Upload Page

IX. FUTURE SCOPE

This Forum can easily support any browser. Various updates may be performed during this forum in returning future. This forum will result in minimizing the gap amongst different users. This forum gives a platform which results in an easy way of communication amongst different people from different areas. Every application has its own merits and demerits. The project has covered almost all the requirements. Further necessities and enhancements will simply be done since the secret writing is principally structured or standard in nature. Changing the present modules or adding new modules will append enhancements. Further enhancements can be made to the application, so that the student community web portal functions very attractive and useful manner than the present one.

X. CONCLUSION

With the proliferation of new computer technologies and Information Technology related innovations, learning and interaction of individuals are gradually taking advantage of these technological developments. Having discussed the structure and some of the features that are present in a forum which makes it a better choice for effective communication among members, it is therefore important that lecturers and students who have regular discussions be advised to require advantage of the pliability that forums provides. With the introduction of forum for student-lecturer interaction, communication between these groups of people become easier as students will be able to participate in the online discussion and contribute their opinion in any material of their choice without being shy or scared of making contributions to topics of dialogue.

XI. ACKNOWLEDGEMENT

We sincerely thank our guide Mrs. Vandana Munde for his guidance and constant support and also for the

stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work. We thank the HOD, Dr. Rajesh Bansode, the Principal, Dr. B. K. Mishra, and the college management for their support.

References

[1] Ankita Singhal, Brij Mohan Kalra," A Literature Review of Online Discussion Forum in e-Learning Scenario", IJARCET,Vol.1,Issue 4,June 2012.

[2] Alabo H. Biriyai, Emmah V. Thomas," Online Discussion Forum: A Tool for Effective Student-Teacher Interaction", IJASR, 2014.

[3] Ravi Seethamraju," Effectiveness of Using Online Discussion Forum for Case Study Analysis", 2014

[4] Wisdom Kwami Takramah, Wisdom Kwasi Atiwoto," Student Database System For Higher Education", American Journal of SEA,2015

[5] A.Oviya, Dr. Kavitha Dr. N. Revathy, I.Sahaya Jaisha, " Discussion forum for students and faculty", IJCSMC, 2019

Machine Learning Based Efficient Rain Prediction

Nadeem Choudhary, Manohar Jaiswar, Ganesh Jha
Information Technology, Thakur College of Engineering and Technology
Mumbai, India
ahmedinadeem@gmail.com, jaiswarmanohar3@gmail.com, 29051198gnj@gmail.com

Abstract— **Forecasting of weather always remained the most important part of human life. Rain forecasting always helped human in natural disaster management which helps in saving precious lives, limited resources and lot of important parameters of existence of livings on the earth. Because of the change in the climate during last 50 years now it has became very important to gather and predict the future events in the weather which helps in flood management agriculture planning and many more disasters . In Traditional methods of weather forecasting, the meteorological departments use verity of tools to gather information about weather and climate, tools include thermometers which measure air temperature, anemometers which gauge wind speeds, and barometers which provide information on air pressure etc. By collecting as much data as possible about the current state of the atmosphere (particularly the temperature, humidity and wind) and using understanding of atmospheric processes (through meteorology) to determine how the atmosphere evolves in the future. However most of the time this method does not provide correct predictions. Additionally weather forecasting using this method requires a good amount of knowledge in the field of meteorology. The proposed system provides a platform which predicts the amount of rainfall can occur in the future based on the historical data available with more accuracy. The methodology used in the proposed system is multiple-linear regression algorithm of machine learning. It requires some independent factors based on which it predicts the dependent factors. The independent factors may include Temperature, Dew point, Sea-level pressure, Visibility, Wind etc and these factors decide the level of precipitation can occur that is rainfall will happen or not? It gives the predictions of rainfall for any point of time for which the temperature, pressure and other required values are known. It does not require any high level knowledge related to meteorology to predict the future events. It also gives graphical representation between various parameters of climate.**

Keywords— Multiple-Linear Regression

I. INTRODUCTION

Climate guaging is an imperious and requesting task that is done by different climate divisions from around the globe so as to advise the ordinary citizens such as ourselves about the climate. It is a complicated system that incorporates different particular fields. This particular objective in this area is convoluted because of the idea of taking choices in a dubious way. An assorted gathering of researchers from everywhere throughout the world has created stochastic climate models with the assistance of Random Number Generator, which can deliver a yield that looks like the climate information, supporting them to fabricate a tremendous dataset with no need of manual recording. India is a nation that comprises of something else than a billion people and over 60% of this populace are reliant upon farming professionally. the uncommon field of precipitation forecast, which is important for all the individuals. Precipitation forecast plays a fundamental in anticipating causalities caused because of catastrophic events. It additionally encourages us to keep up our water assets appropriately. Profoundly precise precipitation forecast is helpful if there should arise an occurrence of substantial precipitation which might cause flood and no precipitation which may cause dry spell to keep up our water assets. In the nation like India where yields are reaped dependent on the storm season. In this way, having propelled information in forecast is significant. A few states are influenced by dry seasons while some with floods. This forecast model can assist us with taking required precautionary measures to spare lives by giving transport and nourishment. Machine learning is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. It has number of algorithms which is used to make a model which can be trained using a dataset. This model can be used to predict the result in the future without any explicit instructions. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

II. RELATED THEORY

Weather plays a very vital role in day to day life, from centuries scientists and meteorologists are trying to predict the change in the events of environment in advance which helps in different aspects of human life, this era is of machine learning which allows us to predict the future events based on the historical data without having a lot of knowledge in that field. There are researches done by numerous researchers on rainfall prediction. We have studied different research papers written by different researchers thus understanding the basic work carried out on this project. We have gone through these papers of different researchers and understood the objectives of these research paper and the research gaps observed. While reading the paper we found that This paper proposes a strategy for selecting the best linear prediction model for Indian monsoon rainfall. This paper results cannot eliminate the possibility that dynamically based, coupled ocean–land–atmospheric models might predict monsoon rainfall with higher skill than linear models. In the next

paper we studied and observed that the present paper analyses the monthly rainfall data. Multiple linear regression is used to predict the average summer-monsoon rainfall using the previous years' data from the corresponding time period. The approach for predicting the rainfall is optimum according to the available data. Although overall prediction error is found to be 26.46%. When reading the next paper we found out that this paper proposed different methods to estimate rainfall. The methods include Autoregressive Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR). However this paper reports a detailed survey on rainfall predictions using different rainfall prediction methods. Although there is a significant limitation in the prediction. The next paper we studied had uses multiple linear regression to predict the monsoon rainfall by using outgoing long wave radiations, global temperatures and sunspots. Although we found out the gap as this paper presents the use of statistical techniques: multiple linear regression method in modeling the rainfall prediction. Though it was efficient but the percentage error was around 15%. When studying this paper we found that This paper presents a case study of using SPSS 13.0 in weather prediction, and the gap we found out was that the present application can forecast an unknown value, using techniques that don't use too many scientific details, thus the precision is not that adequate.

DIFFERENT METHODS OF RAINFALL PREDICTION **:**

MULTIPLE LINEAR REGRESSION

Regression is a statistical measure that attempts to determine the strength of the relationship between one dependent variable usually denoted by Y and a series of other changing variables known as independent variables. Regression model which contain more than two predictor variables are called Multiple Regression Model. Multiple regression model is of the form: Y=b0+b1x1 +b2x2 +b3x3+ b4x4+...e where b0,b1,b2,b3,b4 are regression coefficient e is unexplained portion of dependent variable with zero mean and constant variance. Multiple regression fits a model to predict a dependent (Y) variable from two or more independent (X) variables

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL ARIMA

is used to predict a value in a response time series as a linear combination of its own past values, past errors, and current and past values of other time series. The ARIMA procedure provides a comprehensive set of tools for uni-variate time series model identification, parameter estimation, and forecasting, and it offers great flexibility in the kinds of ARIMA or ARIMAX models that can be analyzed. The ARIMA procedure supports seasonal, subset, and factored ARIMA models; multiple regression analysis with ARMA errors; and rational transfer function models of any complexity. In general, the ARIMA procedure can be subtle as follows :

Step 0) A class of models is formulated assuming certain hypotheses In this step, a general ARIMA formulation is selected to model the rain fall data. This selection is carried out by careful inspection and selection of the main characteristic of the daily rain fall and other meteorological data.The corresponding data are: humidity, air pressure, surface land temperature and wind velocity (corresponding to daily respectively), among others.

Step 1) A model is identified for the observed data. A trial model must be identified for the rain fall data. First, in order to make the underlying process stationary (a more homogeneous mean and variance), a transformation of the original rain fall data and the inclusion offactors of the form trial model must be identified for the rain fall data. First, in order to make the underlying process stationary (a more homogeneous mean and variance), a transformation of the original rain fall data and the inclusion offactors of the form may be necessary. In this step, the checking process can be done using Autocorrelation function (ACF) or unit root test. A further check for lag residual and lag dependent tested from partial ACF.

Step 2) The model parameters are estimated. After the functions of the model have been specified, the parameters of these functions must be estimated. Good estimators of the parameters can be computed by assuming the data are observations of a stationary time series (Step 1). If a Moving Average (MA) pattern is identified then further optimization process needed by using maximum likelihood or least square estimation. A conditional likelihood function is selected in order to get a good starting point to obtain an exact likelihood function. Also, an option to detect and adjust possible unusual observations is selected. As these events are notinitially known, a procedure that detects and minimizes the effect of the outliers is necessary. With this adjustment, a better understanding of the series, a better modeling and estimation, and, finally, a better forecasting performance is achieved.

Step 3) If the hypotheses of the model are validated, go to Step 4, otherwise go to Step 1 to refine the model. In this step, a diagnosis check is used to validate the model assumptions of Step 0. This diagnosis checks if the hypotheses made on the residuals (actual prices minus fitted prices, as estimated in Step 1) are true. Residuals must satisfy the requirements of a white noise process: zero mean, constant variance, uncorrelated process and normal distribution. These requirements can be checked by taking tests for randomness, such as the autocorrelation and partial autocorrelation plots. If the hypotheses on the residuals are validated by tests and plots, then, the model can be used to forecast prices. Otherwise, the residuals contain a certain structure that should be studied to refine the model in Step 1.

Step 4) The model is ready for forecasting. In Step 4, the model from Step 2 can be used to predict future values of daily rainfalldata. Due to this requirement, difficulties may arise because predictions can be less certain as the forecast lead time becomes larger. Based on the natural of data, time series forecasting is suit to short term forecasting (hourly or daily). For a long term period, a structural forecaster is more comply for the situation.

GENETIC ALGORITHM

Genetic algorithms are algorithms that attempt to apply an understanding of the natural evolution in problem-

solving tasks (problem solving). The approach taken by this algorithm is to combine a wide selection of solutions randomly within a population and then evaluate them to get the best solution By doing this process repeatedly, these algorithms simulate the process of evolution as the desired number of generations. This generation will represent improvements on previous population. In the end, we will get the best solutions appropriate to the problems faced. To use a genetic algorithm, solutions to problems represented as a set of genes that make up chromosomes. This chromosome was randomly based coding techniques are used. The entire set of chromosomes is observed representative of the population. Chromosomes will be evolved in several stages iterations called generations. The new generation is obtained By cross breeding techniques (crossover) and mutation (mutation). Crossover includes cutting two pieces of chromosomes based on the desired number of points and then combine half of each chromosome with other couples. While mutations include the replacement value of the gene in a chromosome with the value of other genes from other chromosomes become partner. The chromosomes are then evolved to a suitability criterion (fitness) and the set will be selected the best results while others are ignored. Furthermore, the process repeated until you have a chromosome that has the best fit (best fitness) to be taken as the best solution of the problem. On Genetic Algorithms, the bestsolution search techniques performed simultaneously at a number of solutions known as population. Individuals in a population are referred to as chromosomes. This chromosome is a solution that is shaped symbol. Initial population is built randomly, while the next population is the result of the evolution of chromosomes through iterations called generations. In each generation, the chromosomes will go through an evaluation process using a measurement tool called the fitness function. Fitness value of a chromosome will show the quality of the chromosomes in the population. The next generation is known as the child (offspring) are formed from the combination of two generations of chromosomes that act as the parent (parent) using the crossover operator. Besides crossover operator, a chromosome can also be modified by using mutation operators. The population of the new generation is formed by selecting the fitness value of parent chromosome and the fitness value of the chromosomes of children, and discard the other chromosomes so that the population size (the number of chromosomes in a population) constant. After several generations, the algorithm will converge to the best chromosome.

Genetic Algorithms steps for generating initial weight as follows:

1.Create an initial population randomly of meteorological data.

2. Evaluate each individual in the population.

3.Generate new population using genetic operations.

4.Determine the final result at the time of termination criteria.

ADAPTIVE SPLINES THRESHOLD AUTOREGRESSIVE (ASTAR)

Modelling In modelling ASTAR several software are used and integrated to process the ASTAR result, i.e.Microsoft Excel, SPSS 16 and MARS 2.0 are the software for ASTAR planning system.Rain fall forecasting, as response variable (Y), Input variable, as predictor variable (X), is wind speed, humidity and temperature withX1, X2, and X3 respectively. All of predictor variables are applied to attain the best model of rainfall forecasting.The significant variable, influenced the next day condition with importance variable, is processed using MARS 2.0 Software. Function Base A Basis Function is distance between sequence knots. In ASTAR, Basis Function is a set of function to describe information that consist of one or two variables. Max(0, x − t) or Min (0, t − x) is Basis Function value with t as a value to illustrate knot position and x as predictor variable. Every 1 knot will produce a couple of Basis FASTAR Methods as data analysis technique to find the best model from a set of data. It is using past and present data to predict the short-term forecasting.

Modelling Stage of ASTAR

1.Determine maximum Basis Function, maximum interaction numbers and minimum observation numbers between knots.

2.Forward Stepwise Processing to obtain maximum number of Basis Function using MARS 2.0

3.Backward Stepwise Processing to obtain Basis Function numbers from forward stepwise by minimizing the least GCV (Generalized Cross Validation) value.

4.Knots selection using forward and backward algorithm. 5.Estimating the coefficient of chosen Basis Function as a stage of response variable (Y) prediction (Y) to predictor variable (X).

SUPPORT VECTOR MACHINE (SVM)

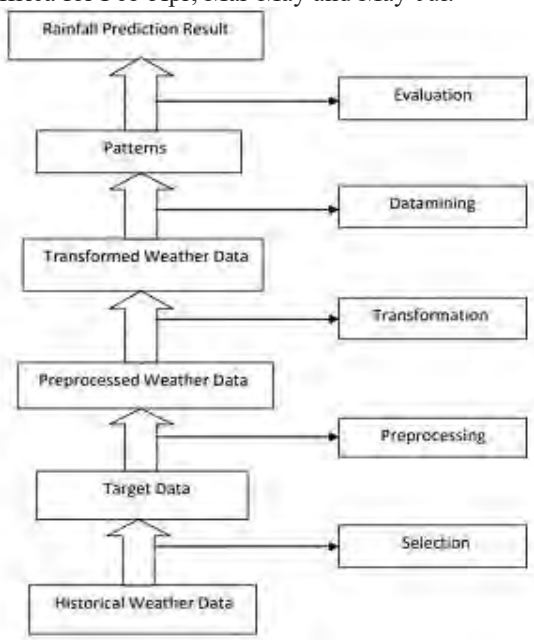
A Support Vector Machine (SVM) is a computer algorithm that learns by example to find the best function of classifier hyperplane to separate the two classes in the Input space.The SVM analyzed two kinds of data, i.e. linearly and non- linearly separable data .The example of linearly separated data is shown in fig. below. Best hyperplane betweentwo classes can be foundby measuring the hyperplanemargin and find out the maximum points.Margin is defined as the distance between hyperplane and the closest pattern of each class, which is called support vector.The best hyperplane is defined by the following equation f(x) = w x+ b(1) T Where x refers to a training pattern, w is referred to as the weight vector and b as the bias termSupport Vector Machine is one of the important category of perceptrons and radial basis function networks, support vector machines can be used for pattern classification and nonlinear regression. Support Vector Machines (SVMs) developed by Vapnik and his co-workers has been used for supervised learning due to – (i) Better generalization performance than other NN models (ii) Solution of SVM is unique, optimal and absent from local minima as it uses linearly constrained quadratic programming problem (iii) Applicability to non- vectorial data (Strings and Graphs) and (iv) Few parameters are required for tuning the learning m/c. Kernel Methods are a set of algorithms from statistical

learning which include the SVM for classification and regression, Kernel PCA, Kernel based clustering, feature selection, and dimensionality reduction etc . SVM is found to be a significant technique to solve many classifications problem in the last couple of years. Very few researchers of this field used this technique for rainfall prediction and got satisfactory result.

WEATHER RESEARCH AND FORECASTING MODEL The Weather Research and Forecasting (WRF) model is a numerical weather Prediction (NWP) and atmospheric simulation system desined for both research and operational applications. The development of WRF has been a multi-agency effort to build a next-generation forecast model and data assimilation system to advance the understanding and prediction of weather and accelerate the transfer of research advances into operations. The geogrid defines model domains and interpolates static geographical data to the grids. ungrib extracts meteorological fields from GRID formatted files. The metagrid horizontally interpolates the meteorological fields extracted by by ungrib to the model grids defined by geogrid. Each of the WPS programs reads parameters from a common name list file, as shown in the figure. This name list file has separate name list records for each of the programs and a shared namelist record, which defines parameters that are used by more than one WPS program The ungrib program reads GRIB files degribs the data, and writes the data in a simple format, called the intermediate format.GRIB (Gridded Binary or General Regularly distributed Information in Binary form) is a mathematically concise data format commonly used in meteorology to store historical and forecast weather data.

Seasonal Climate Forecasting

The CGCM is run by the BoM out for 9 months every day. Forecast products are generated from dynamical model output using data analysis software. The resulting derived forecast products are persisted in self describing files with additional metadata to support the clients that deliver the outlooks. Forecast data is exposed via a data server. Scheduled processes access and reformat the data for SCOPIC (Seasonal climate outlooks for pacific island Countries) access. Custom web services use the data server's interface to the forecast data to provide maps, data, and line plots. The Pacific Adaptation Strategy Assistance Program (PASAP) Portal consumes the outputs of the custom web services, and displays model based outlooks as overlays on dynamical maps and standard plots. The high predictability of seasonal climate in the tropical Pacific provides opportunities for using seasonal forecasts to improve the resilience of climate sensitive sectors throughout the region. An enhanced probability for below-normal precipitation is forecast for Feb-Apr and Mar-May in parts of Indonesia, northern Australia, and for Feb-Apr in part of southeast Africa. Below-normal is also favored for Feb-Apr through Apr-Jun in the Philippines, for Feb-Apr in scattered portions of interior Asia, and for May-Jul in central Chile and coastal northeast Brazil. Enhanced chances for above-normal precipitation are predicted for Feb-Apr in coastal northeast Brazil, northeastern Scandinavia, and northeast Asia, and for Feb-Apr and Mar-May in far northern North America. A tilt of the odds toward above-normal is also forecast in eastern equatorial Africa for Feb-Apr, Mar-May and May-Jul.



Since 2004 the Pacific Island-Climate Prediction Project (PI-CPP) managed by the Australian Bureau of Meteorology (BoM) has built seasonal prediction capabilities within National Meteorological Services (NMS) of Pacific Island countries through the development and provision of decision support software

GENERAL DATA MINING RAINFALL PREDICTION MODEL: In general data mining prediction model first we collect the historical weather data. Data were collected from Indian metrological department pune. the collected data consist of different features including daily dew point temperature (Celsius), relative humidity, wind speed (KM/H), Station level pressure, Mean sea level, wind speed, pressure and rainfall observation. Creating a target data set selecting a data set or focusing on a subset of variables or data samples on which discovery is to be performed. Then important step in the data mining is data preprocessing. One of the challenges that face the knowledge discovery process in meteorological data is poor data quality. For this reason we try to prepare our data carefully to obtain accurate and correct results. First we choose the most related attributes to our mining task. purpose we neglect the wind direction. Then we remove the missing value records. In our data we have little missing, because we are working with weather data. Then finding useful features to represent the data depending on the goal of the task.

After preprocessing and transforming the weather data choosing the data mining task i.e. classification, regression and decision tree. Then applying different data mining techniques i.e. K-NN, Naïve Bayesian, Multiple Regression and ID3 on weather data set and makes the rainfall prediction i.e. Rainfall Category or No Rainfall Category.

III. PROPOSED METHODOLOGY

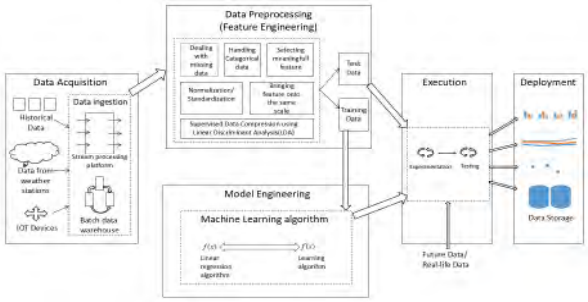


Fig. Proposed Methodology Architecture

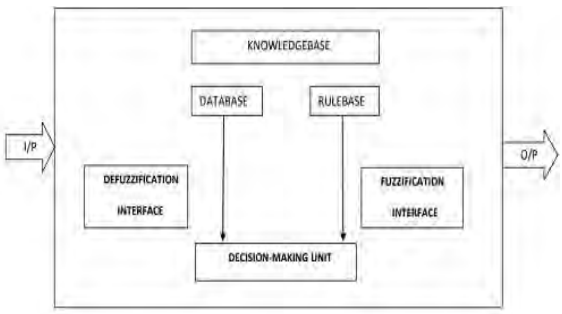
In this project we have tried to predict the rainfall using machine learning concept. The methodology is undertaken into different steps , basically three stages. Data Acquisition, Data pre-processing and Data Execution. In the first step we try to acquire as many data available which are relevant to our project from different resources. To build a prediction model we need an efficient dataset to train our model for the required output, so to have efficient and relevant data is very necessary before developing a prediction model. We can have different resources for these data, such as historical data which are already recorded and stored from past, or data which are real time like weather factors such as temperature , pressure etc. Also sometime the data is acquired from different sensors, relative to the system. As all these data are acquired from different sources, there is a possibility of this data being in unstructured format, also it maybe a noisy too. Thus to make this data usable and suitable for the model, we need to preprocess the data. We have to remove any anomaly or discrepancies from these datasets to make it system dependent. Also a machine learning model will be implemented in which this acquired data will be feeded to, for training the model and then performing a testing of the accuracy of this model with the help of remaining data.

WEATHER RESEARCH AND FORECASTING MODEL The Weather Research and Forecasting (WRF) model is a numerical weather Prediction (NWP) and atmospheric simulation system desined for both research and operational applications. The development of WRF has been a multi-agency effort to build a next-generation forecast model and data assimilation system to advance the understanding and prediction of weather and accelerate the transfer of research advances into operations. The geogrid defines model domains and interpolates static geographical data to the grids. ungrib extracts meteorological fields from GRID formatted files.The metagrid horizontally interpolates the meteorological fields extracted by by ungrib tothe model grids defined by geogrid. Each of the WPS programs reads parameters from a common namelist file, as shown in the figure.This namelist file has separate namelist records for each of the programs and a shared namelist record, which defines parameters that are used by more than one WPS program The ungrib program reads GRIB files degribs the data, and writes the data in a simple format, called the intermediate format.GRIB (Gridded Binary or General Regularlydistributed Information in Binary form) is a mathematically concise data format commonly used in meteorology to store historical and forecast weather data.

GLOBAL DATA FORECAST SYSTEM

A new Global Forecast System (GFS) has been implemented at Northern Hemisphere Analysis Center of IMD on High Power Computing Systems (HPCS). The new GFS is running in experimental real-time model since 15th January 2010. This new higher resolution global forecast model. The GFS at IMD Delhi involves 4 steps as given below: Steps 1 - Data Decoding and Quality Control: First step of the forecast system is data decoding. It runs 48 times in a day on half-hourly basis, as soon as GTS data files are updated at regional telecom hub (RTH) of global telecom system (GTS) at IMD New Delhi. Steps 2- Preprocessing of data: (PREPBUFR): Runs 4 times a day at 0000, 0600, 1200 & 1800 UTC. Step 3 - Global Data Assimilation (GDAS) cycle: The Global Data Assimilation cycle runs 4 times a day (00, 06, 12 and 18 UTC). The assimilation system is a global 3-dimensional variational technique, based on NCEP's Grid Point Statistical Interpolation (GSI) scheme, which is the next generation of Spectral Statistical Interpolation (SSI). Step 4 – Forecast Integration for 7 days: The analysis and forecast for 7 days is performed using the HPCS installed in IMD Delhi. One GDAS cycle and seven dayforecast (168 hour) run takes about 30 minutes.

FUZZY LOGIC (FUZZY) Fuzzy Logic is a type of reasoning based on the recognition that logical statements are not only true or false (white or black areas of probability) but can also range from "almost certain" to "very unlikely".Fuzzy logic has proven to be particularly useful in expert system applications. Fuzzy inference system is shown in diagram below. They are composed of five conventional blocks a rule-base containing a number of fuzzy if-then rules, a database which defines the membership functions of the fuzzy sets used in the fuzzy rules, a decision making unit which performs the inference operations on the rules, a fuzzification interface which transform the crisp inputs into degrees of match with linguistic values, a defuzzification interface which transform the fuzzy results of the inference into a crisp output.



IV. RESULT&DISCUSSION

The Amount Of Rain That Can Happen In Near Future Can Be Predicted Using Historical Data Available On Kaggle. The Expected Result Is Represented In The Graphical Format.

V. APPLICATION

Rainfall plays a very critical role in every aspects of life. The rainfall can either boost the society growth or destroy the society. It is thus very important to consider the rainfall as the main aspect for deciding other event. India being an agricultural country, extensively depends on Rainfall for its prosperous growth. Thus this project a wide range of spectrum for application in real life. From stepping out of our home, to economy of this country it mostly depends on rainfall, thus it is becomes very important to know wether there will be a rain or not, or how much rainfall is estimated. Thus this project has various application scope in the real-life.

VI. CONCLUSION

In this project, we've got suggested an analysis of the soil information using Decision Tree algorithms and prediction technique. By using Decision tree algorithm, we recommend farmers about their field soil quality and suggest crops which are suitable to grow in that soil. Agriculture is the utmost important area especially in the mellowing country like India. Use of information technology in agriculture can change the scenario of decision making and framers can yield in a better way. For decision making on overall issues related to agriculture field; data mining plays a vital role. The survey discussed the role of data mining in terms of agriculture field. We have also discussed various types of soils, several data mining techniques in agriculture and soil containment.

References

[1] G. Bala Sai Tarun, J.V. Sriram, K. Sairam, K. Teja Sreenivas, M.V.B.T. Santhi "Rainfall prediction using Machine Learning Techniques.

[2] S. Monira Sumi, M. Faisal Zaman "A Rainfall Forecasting Method Using Machine Learning Models and its Application to the Fukuoka City Case"

[3] E. B. Abrahamsen, O. M. Brastein, B. Lie “Machine Learning in Python for Weather Forecast based on Freely Available Weather Data”

[4] Machine Learning Model on Rainfall - A Predicted Approach for Bangladesh.

[5] Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee" by A H M Jakaria, MD Mosharaf Hossain, MD Ashiqur Rahman.

[6] N. Chantasut, C. Charoenjit, and C. Tanprasert, “Predictive mining of rainfall predictions using artificial neural networks for Chao Phraya River,” 4th Int Conf. of the Asian Federation of Inform. Technology in Agriculture and the 2nd World Congr. on Comput. in Agriculture and Natural Resources, Bangkok, Thailand, pp. 117-122, 2004. [7] V. K. Somvanshi, O. P. Pandey, P. K. Agrawal, N.V.Kalanker1, M.Ravi Prakash, and Ramesh Chand, “Modeling and prediction of rainfall using Artificial Neural Network and ARIMA techniques,” J. Ind. Geophys. Union, vol. 10, no. 2, pp. 141-151, 2006.

[8] S. Chattopadhyay, “Anticipation of summer monsoon rainfall over India by Artificial Neural Network with Conjugate Gradient Descent Learning,” arXiv preprint nlin/0611010, pp. 2-14, 2006.

[9] S. Chattopadhyay and M. Chattopadhyay, “A Soft Computing technique in rainfall forecasting,” Int. Conf. on IT, HIT, pp. 19-21, 2007.

[10] S. Chattopadhyay and G. Chattopadhyay, “Comparative study among different neural net learning algorithms applied to rainfall time series, "Meteorological applicat., vol. 15, no. 2, pp. 273-280, 2008.

[11] P.Guhathakurta, “Long lead monsoon rainfall prediction using deterministic Artificial Neural Network model,” Meteorology and Atmospheric Physics 101, pp. 93-108, 2008.

[12] C. L. Wu, K. W. Chau, and C. Fan, “Prediction of rainfall time series using Modular Artificial Neural Networks coupled with datapreprocessing techniques," J. of hydrology, vol. 389, no. 1, pp. 146- 167, 2010.

[13] K. K. Htike and O. O. Khalifa, “Rainfall forecasting models using Focused Time-Delay Neural Networks,” Comput. and Commun. Eng. (ICCCE), Int. Conf. on IEEE, 2010.

[14] S. Kannan , Subimal Ghosh, “Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output”, Springer-Verlag, July- 2010.

[15] M.Kannan, S.Prabhakaran, P.Ramachandran, “Rainfall Forecasting Using Data Mining Technique”, International Journal of Engineering and Technology Vol.2 (6), 397-401, 2010.

[16] G. Geetha and R. S. Selvaraj, “Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model,” Int. J. of Eng. Sci. and Technology, vol. 3, no. 1, pp. 211 213, 2011.

[17] M. A. Sharma and J. B. Singh, “Comparative Study of rainfall forecasting models,” New York Sci. J., pp. 115-120, 2011.

[18] J. Abbot and J. Marohasy, “Application of Artificial Neural Networks to rainfall forecasting in Queensland, Australia,” Advances in Atmospheric Sci.,vol. 29, no. 4, pp. 717-730, 2012.

[19] A. Kumar, A. Kumar, R. Ranjan, and S. Kumar, “A rainfall prediction model using artificial neural network,” Control and Syst. Graduate Research Colloq. (ICSGRC), pp. 82-87, 2012.

[20] R. R. Deshpande, “On the rainfall time series prediction using Multilayer Perceptron Artificial Neural Network,” Int. J. of Emerging Technology and Advanced Eng., vol. 2, no. 1, pp. 148-153, 2012.

[21] Soo-Yeon Ji, Sharad Sharma, Byunggu Yu, Dong Hyun Jeong, “Designing a Rule-Based Hourly Rainfall Prediction Model”, IEEE IRI 2012, August – 2012.

[22] G. Shrivastava, S. Karmakar, and M. K. Kowar, “BPN model for long range forecast of monsoon rainfall over a very small geographical region and its verification for 2012,” Geofizika, vol. 30, no. 2, pp. 155-172, 2013.

[23] C. L Wu and K. W. Chau, “Prediction of rainfall time series using modular soft computing methods,” Eng. Applicat. of Artificial Intell., vol. 26, no. 3, pp. 997-1007, 2013.

[24] S. K. Nanda, D. P. Tripathy, S. K. Nayak, and S. Mohapatra, “Prediction of rainfall in India using Artificial Neural Network (ANN) models,” Int. J. of Intell. Syst. and Applicat., vol. 5, no. 12, pp. 1-22, 2013.

[25] A. R. Naik and S. K. Pathan, “Indian monsoon rainfall Classification and Prediction using Robust Back Propagation Artificial Neural Network,” Int. J. of Emerging Technology and Advanced Eng., vol. 3, no. 11, pp. 99-101, 2013.

[26] Priya, Shilpi, Vashistha and V. Singh, “Time Series Analysis of Forecasting Indian Rainfall,” Int. J. of Innovations & Advancement in Comput. Sci., vol. 3, no. 1, pp. 66-69, 2014.

[27] V. K. Dabhi and S. Chaudhary, “Hybrid Wavelet-Postfix-GP model for rainfall prediction of Anand region of India,” Advances in Artificial Intell., pp. 1-11, 2014.

[28] Pinky Saikia Dutta, Hitesh Tahbilder, “Prediction Of Rainfall Using Data mining Technique Over Assam”, Indian Journal of Computer Science and Engineering (IJCSE), Vol. 5 No.2 Apr May 2014.

[29] Pankratz, A., Forecasting With Univariate Box-Jenkins Models Concepts and Cases.John Wiley Sons, Inc. New York, pp: 414.

[30] M. T.Hagan, H.B.Demuth and M.Beale, Neural Network Design Thomson Asia Pte. Ltd , Singapore, 2002.

Career Guidance using Machine Learning

Mr. Dishant Singh, Mr. Hardik Vishwakarma, Mr. Umesh Pal, Dr. Sandip Bankar
(TCET-IT, Student) Mumbai, India
dishantsingh1998@gmail.com, hardikrv007@gmail.com, umesh983311@gmail.com

Abstract – **A career guidance system where students can see various career opportunities, the system shows different fields available after the 10th, 12th for graduation and after graduation fields are available. It also lists several universities available where students can search for universities for their courses. Subsequently, the system allows users to perform a test. The tests have two levels. The first level has questions that help us identify which field the student is interested in and, depending on this, the second level will ask questions from their field of interest. The first level has the same questions for all standards, since it is not based on knowledge but on interests. The second level questions will come from the field in which the student is most interested, together with the questions from other fields. After successfully completing both levels, the student will receive the results along with a detailed explanation and advice from us. The explanation will indicate why we believe a certain career is better for him and why not the other.**

I. INTRODUCTION

Choosing a right career for our self has been one of the most important decision of our lives from a very early age. This decision always keeps tricking our mind as to what our interest lies in. The tension starts to mount gradually as we grow up and realise that we have reached a point where we need to decide what has to be done next in life, but are still confused with the idea.

Career guidance system is designed by keeping in mind the seriousness of one's career. Our website is a relief for the students as it will clear a lot of their confusion and save time. Our system will guide the students after ssc,hsc and graduation who are looking for best career opportunities.

This system is designed in such a way that students will be able to easily rectify their area of interest and move further with the process of selecting a career for themselves. Student will have to first register with our website. Then they will have to login to the system with their id which will be given to them by the system. The first step will be that the student will have to select his/her standard to move further.

Student will then enter his/her data and this data will analysed by the system to give accurate results. This system will be connected to the database where all the information of the student will be stored.

Career guidance system will work as a guide for students and help them in analysing their interest and in making the right choice for the future. This system will show various career opportunities to the students based on the data entered by them.

Students can see suitable streams available after 12th or graduation. Students will be able to pick or select only one stream from science /commerce/arts. Based on the data entered by the student, their area of interest will be analysed by the system.

An online aptitude test will be taken to help the student in making a suitable choice. This aptitude test will be available for the students to make choices based on their knowledge. The result of the test will be given to the student and based on the marks scored by the student, career choice will be given to the student. E-books will be made available via hyperlinks for students who are looking for more information on certain fields. The system will generate a list of available fields and courses for the students based on their academics and interest. With the help of the guidance provided by this system, career selection will become easier for the student. Best colleges/universities will be marked based on their score. Students can also view this list of colleges based on the courses selected by them. This system will help students in selecting the right streams based on their performance. With the help of our system, confusion of selecting the right path for future will be solved and time and cost will be saved overall.

II. Literature Survey

The traditional approach to career guidanceis a manual method that is ineffective and inefficient. The electronic approach provides an effective and efficient career guidance. The intelligent system uses direct parameters from the students, such as the combination of preferred scientific subjects, the result of the analysis of the inventory of professional interests and the result of the intelligent quotient test for the professional recommendation. The web-based intelligent system was designed and implemented with the principle of a rule-based expert system using the fasttree algorithm, the client / front-end pages were designed using the "Bootstrap 3" frontend

framework containing: HTML5, CSS3 and JavaScript SSMS 2018 was used for the back end.

III. METHODOLOGY

The methodology that we chose for our project is Agile. While waterfall model is a tried and tested framework, it is not suitable for our project, due to the changeable nature and small tea size. Agile allows us to be flexible, manage changing requirements, manage the ever-increasing scope as well as get consumer perspective.

Student :
This module provides functionality for student. Applicants/students can give online Aptitude test System will automatically predict career option and courses for student based rest result. the applicant cans also browse through courses and colleges available. A student registration page will be created for the students who are looking for career guidance. Students will have to register themselves on the website in order to avail the facilities. Registered students will get a login id and password. They can use the website by logging in using the generated login id and password.
displayed.

Available courses: Students looking for certain courses can view it on the website. Also, the available courses after 10th, 12th or graduation will be displayed. Courses available for the student based on the marks will also be displayed. **Streams/fields:** Student can directly select courses according to their interest or available streams will be listed to the students based on their performance.

Suggestion: Streams will be suggested by the system to the students based on their performance

IV. PROJECT PROTOTYPE

PROTOTYPE DEVELOPMENT

The traditional approach to career guidance is a manual method that is ineffective and inefficient. The electronic approach provides an effective and efficient career guidance. The intelligent system uses direct parameters from the students, such as the combination of preferred scientific subjects, the result of the analysis of the inventory of professional interests and the result of the intelligent quotient test for the professional recommendation. The web-based intelligent system was designed and implemented with the principle of a rule-based expert system using the fasttree algorithm, the client / front-end pages were designed using the "Bootstrap 3" frontend framework containing HTML5, CSS3 and JavaScript SSMS 2018 was used for the back end.

3.2 Product Backlog or Sprint Backlog

The backlog of the project is given as follows:

To do

- Make GUI for fringe features
- Add a option for Sender to send type of data
- Add an archiving feature
- Providing Estimating transfer time
- Make a checklist database for users
  Ongoing
- Information catalogue building
- Dataset training
- GUI building
- Procurement of data for higher accuracy Done
- Analysis

Design

- Modelling
- Literature survey
- Budgeting and scheduling
  Feasibility analysis

3.3 Project planning (Resources, Tools used, etc.)

IT projects require resources in terms of money, time, human resources, infrastructure and technology, both hardware and software. Resources are not just a mean, but also an approximation of constraints. Project planning is essential to managing the scope, schedule and budget of the project. For this, we used tools such as MS Excel, MS PowerPoint, online MS project as well as various modelling tools, such as draw.io. Thus, we made it a necessity to search for user consensus before we planned for features to be built in our project. This was done by researching search

interest on search engines, visiting forums pertaining to QR code. We divided our project into various phases and sub-phases, and allocated date ranges from a week to 3 weeks to every sub-phase. This was done using the timeline chart feature of online MS project. We then used a Gantt chart to model schedule dependencies and fine tune the scheduling. The Gantt chart was created using the Gantt chart feature available in MS Excel. Using these two, we further planned our sub-phases.

The modelling included charts and diagrams such as UML diagram, GUI modelling, dependencies and other diagrams. These were done using free tools like draw.io. Drawing these models helped to incisively assess our requirements and features.

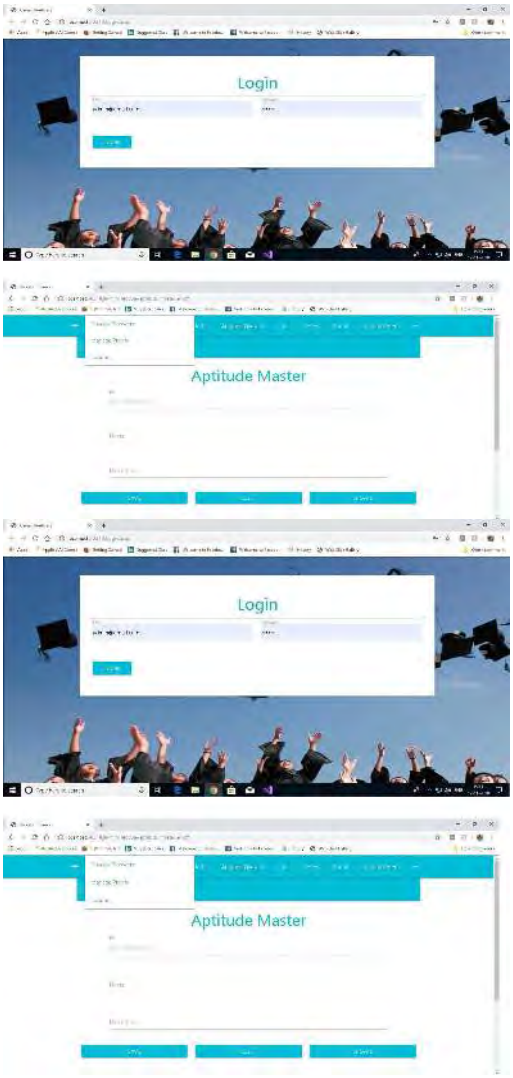
V. FUTURE SCOPE

In the knowledge economy and society, the trend towards aligning higher education with the needs of the working world has received considerable attention, turning concepts such as
'evaluation', 'accountability', 'knowledge society', 'knowledge economy', 'employability', and 'internationalisation' or 'globalisation' into some of the main recurring dimensions of the debate about the relationships between higher education and the world of work.In this context, a diverse range of support services geared towards helping students with the transition to employment has become increasingly common among higher education providers in recent decades. As graduation from higher education does not guarantee the transition to employment or professional success, higher education is considered an imperfect source of training for the world of work. The transition from one sphere to the next is a process influenced by a wide range of internal and external factors affecting professional development. In this context, a diverse rangeof support services geared towards helping students with the transition to employment has become increasingly common among higher education providers in recent decades. As graduation from higher education does not guarantee the transition to employment or professional success, higher education is considered an imperfect source of training for the world of work. The transition from one sphere to the next is a process influenced by a wide range of internal and external factors affecting professional development.

VI. CONCLUSION

The career guidance system is an innovative idea. The opportunities offered by this electronic medium are immense and many students can take advantage of this medium to choose a career more suited to their abilities. In today's competitive and technological world, with countless options available, the student is generally confused in choosing the right or most appropriate career. The world these days is moving towards "information flows". The information is released to the user instead of the user learning the information. In light of the above, it is believed that the proposed system has the ability to connect with different students and help them connect with the most appropriate career.

VII. RESULTS



References

[1] Essaid EL HAJI, Abdellah AZMANI, Mohamed EL HARZLI(2014), "Multi-expert system design for educational and career guidance: an approach based on a multi-agent system and ontology," Department of Computer Science, LIST Workeratory, Faculty of Science and Technology.

[2] S.Saraswathi, "Design of an online expert system for career guidance", Department of Computer Science, LIST Workeratory,Faculty of Science and Technology.

[3] IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2, September 2014 ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784. [Accessed 09 October 2015].

4] Ankit, M., Ashutosh, S., Sunil, K. S., Pardeep, K., and Durg, S. C. (2014). Decision Support System for Determining: Right Education

Career Choice. Elsevier, Vol. 4, No.1, pp. 8 – 174.Xiuhua, J, 2013.

[5] FFMPEG'. Fourth World Conference on Software Engineering. I Chang,Y Liao Volume No (6) 2014 C, CWirCommunications and Mobile Computing Conference (IWCMC), 2015 International, pp.665-667. Society (SCCC), 2015.

[6] Chung, Y, Sewaiwar, A, Vikramaditya, V & Tiwari, S 2015, 'Novel user allocation scheme for full duplex multiuser bidirectiona network', Optics Communications, vol. 339, pp. 153-156. Clifford, P 2016.

[7] L.Balaguru Swamy and Charles Kingston GPIO Magazine on Machine Learning in modern 21st century .IEEE 2018 Volume No 9 Section II ,2018

Stock Market Prediction Using Neural Network

Bhavesh Chaudhary, Vinayak Poojary, Sudhir Dhekane , Aditya Mohite,
Information Technology
Thakur College of Engineering and Technology
Mumbai, India
bhaveshtpchaudhary@gmail.com, vinayakpoojary8@gmail.com, sudhirdhekane1987@thakureducation.org, aditm981126@gmail.com

Abstract - **A stock market is a public market for the trading of company stock. It is an organized set-up with a regulatory body and the members who trade in shares are registered with the stock market and regulatory body SEBI. Since stock market data are highly time-variant and are normally in a nonlinear pattern, predicting the future price of a stock is highly challenging. Prediction provides knowledgeable information regarding the current status of the stock price movement. Many researches have carried out for predicting stock market price via various data mining techniques. This focuses on use of Artificial Neural Network techniques to predict the stock price. The past data of the selected stock will be used for building and training the models. The results from the model will be used for comparison with the real data to obtain the accuracy of the model.**

Predicting financial market changes is an important issue for time series analysis, receiving an increasing attention in past years.

The mean square error (MSE) measure was used to evaluate the performances of proposed model. The experimental output lead to the conclusion that the proposed model can be successfully used as an alternative method to standard statistical techniques for financial time series forecasting.

I. INTRODUCTION

In modern world financial time series prediction is an important subject for many financial analysts and researchers as accurate forecasting of different financial applications play a key role in investment decision making. Stock market prediction is one of the most difficult tasks of time series analysis since the financial markets are influenced by many external social-psychological and economic factors. Efficient market hypothesis states that stock price movements do not follow any patterns or trends, and it is practically impossible to predict the future price movements based on the historical data. However, financial time series are generally non-stationary, complicated and noisy, it is possible to design mechanisms for prediction of financial markets. Technical analysis with statistical and machine learning techniques have been applied to this area in order to develop some strategies and methods to be helpful for financial time series forecasting. The statistical methods include Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM), area of stock prediction, feature selection plays a significant role in forecasting accuracy and efficiency.

RNN:- Recurrent Neural Network(RNN) are a type of Neural Network where the output from previous step are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence. RNN have a "memory" which remembers all information about what has been calculated.

RNN converts the independent activations into dependent activations by providing the same weights and biases to all the layers, thus reducing the complexity of increasing parameters and memorizing each previous outputs by giving each output as input to the next hidden layer. Hence these three layers can be joined together such that the weights and bias of all the hidden layers is the same, into a single recurrent layer.

II. OBJECTIVE AND SCOPE

The aim of the project is to examine a forecasting technique to predict future stock returns based on past returns We do this by applying Neural network methods like RNN (Recurrent neural network) and LSTM(long short term memory) for stock price forecasting by interpreting the seemingly chaotic market data.

The stock market prediction task is interesting as well as divides researchers and academics into two groups those who believe that we can devise comes mechanism to predict the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction.

We are going to predict market price (index) of each individual company's share with help of historical data of that particular company,Similarly we can implement many of the company's share price individually and merge them in to a particular User interface, also we would check the accuracy on with the help of historical data which will not be present In trained.

III. PROBLEM DEFINITION

The uncertainity of the stock market and the unstable growth of any stock exchange creates challenges among investors to choose the right stock as per the market behaviour, to overcome these challenges stock market prediction models play a very critical role in influencing the investing behaviours of investors. There are a lot of

complicated financial indicators and also the fluctuation of the stock market is not very stable. But as the technology is getting advanced, the opportunity to gain a steady fortune from the stock market is increased and it also helps experts to find out the most informative indicators to make a better prediction. The prediction of the market value is of great importance to help in maximizing the profit of stock option purchase while keeping the risk.

The stock market prediction task is interesting as well as divides researchers and academics into two groups those who believe that we can devise comes mechanism to predict the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction.

We are going to implement a system in such a way so that all the problems which previously have been occured will be solved by using LSTM and RNN techniques of neural networks by taking a particular stock company data and train them as time series analysis based solution.

IV. METHODOLOGY

The prediction of the market value is of great importance to help in maximizing the profit of stock option purchase while keeping the risk low. Recurrent neural networks (RNN) have proved one of the most powerful models for processing sequential data.

Long Short-Term memory is one of the most successful RNNs architectures. LSTM introduces the memory cell, a unit of computation that replaces traditional artificial neurons in the hidden layer of the network. With these memory cells, networks are able to effectively associate memories and input remote in time, hence suit to grasp the structure of data dynamically over time with high prediction capacity.

In our LSTM model for stock prediction, one sequence was defined as a sequential collection of the daily dataset of any single stock in a fixed time period (N days). The daily dataset describes the performance of the stock with sequence learning features like closing price, trade volume on one particular day in these N days.

Our model is composed of (1) a single input layer with the number of memory cells as that of the sequence learning features one sequence may hold, followed by 2) multiple LSTM layers and (3) a dense layer and (4) a single output layer.

This section we will discuss the methodology of our system. Our system consists of several stages which are as follows:-

- Stage 1: Raw Data: In this stage, the historical stock data is collected from Google Finance and this historical data is used for the prediction of future stock prices.
- Stage 2: Data Pre-processing:

The pre-processing stage involves,

a) ***Data discretization***: It is a part of data reduction but with particular importance, especially for numerical data

b) ***Data transformation***: Normalization.

c) ***Data cleaning***: Fill in missing values.

d) ***Data integration***: Integration of data files.

After the dataset is transformed into a clean dataset, the dataset is divided into training and testing sets so as to evaluate. Here, the training values are taken as the more recent values. Testing data is kept as 5-10 percent of the total dataset.

Stage 3: Feature Extraction: In this layer, only the features which are to be fed to the neural network are chosen. We will choose the feature from Date, open, high, low, close, and volume.

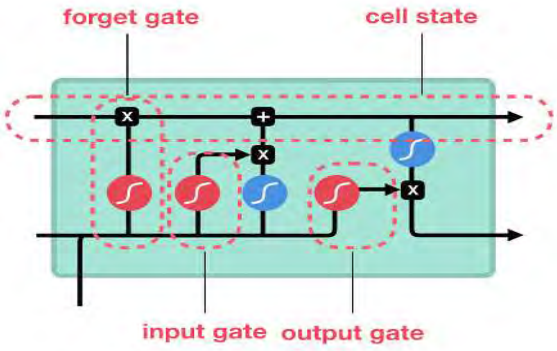
Stage 4: Training Neural Network: In this stage, the data is fed to the neural network and trained for prediction assigning random biases and weights. Our LSTM model is composed of a sequential input layer followed by 2 LSTM layers and dense layer with ReLU activation and then finally a dense output layer.

Stage 5: Output Generation: In this layer, the output value generated by the output layer of the RNN is compared with the target value. The error or the difference between the target and the obtained output value is minimized by using back propagation algorithm which adjusts the weights and the biases of the network.

LSTM recurrent unit tries to “remember” all the past knowledge that the network is seen so far and to “forget” irrelevant data. This is done by introducing different activation function layers called “gates” for different purposes. Each LSTM recurrent unit also maintains a vector called the Internal Cell State which conceptually describes the information that was chosen to be retained by the previous LSTM recurrent unit. A Long Short Term Memory Network consists of four different gates for different purposes as described below:-

Forget Gate(f): It determines to what extent to forget the previous data.

Input Gate(i): It determines the extent of information to be written onto the Internal Cell State.



Output Gate(o): It determines what output(next Hidden State) to generate from the current Internal Cell State.

The first step in our LSTM is to decide what information we're going to throw away from the cell state. This decision is made by a sigmoid layer called the "forget gate layer." The next step is to decide what new information we're going to store in the cell state. This has two parts. First, a sigmoid layer called the "input gate layer" decides which values we'll update. Next, a tanh layer creates a vector of new candidate values that could be added to the state. In the next step, we'll combine these two to create an update to the state.It's now time to update the old cell state into the new cell state.Finally, we need to decide what we're going to output. This output will be based on our cell state, but will be a filtered version. First, we run a sigmoid layer which decides what parts of the cell state we're going to output. Then, we put the cell state through tanh (to push the values to be between −1 and 1) and multiply it by the output of the sigmoid gate, so that we only output the parts we decided to.

V. EXPECTED RESULT

Stock market Prediction techniques will obvious make an improvement of Commercial sector of Business. With the use of LSTM (Long Short Term memory) algorithm of machine learning we analyzed the result of predictive stats of a particular Company's Stock. with the help of historical data such as Date, Open, close ,Volume, high, low,respectively. And with the help of these attribute we predicted the similar result of previous day stock which was not considered for dataset.

VI. CONCLUSION

This prediction model will help to develop neural network as another forecasting tool for highly volatile Indian market. In this paper, we attempted to find an optimal architecture of the neural network to predict the direction of the CNX S&P Nifty 50 Index with a high level of accuracy.

This method has considered most of the issues and critical factors for designing the neural network model and has tested the performance of each of the structures at various historic periods of the trading sessions of the Indices. A rigorous trial and error method is employed in selecting each of the features of the network structure. A three layer feed-forward back propagation neural network with 10 input neurons, a hidden layer with 5 neurons and 1 output neuron with tan sigmoid and linear transfer function in the hidden and output layer respectively is considered as the optimal network structure.

The structure is tested with the most recent trading data set. The input data used in the model is the preprocessed historic closing values of the index. The model predicts the next trading days closing value of the index. The training data set comprises of the historic data previous to the prediction day of prediction.

References

[1]. Olivier C., Blaise Pascal University: "Neural network modelling for stock movement prediction, state of art". 2007
[2]. Leng, X. and Miller, H.-G. : "Input dimension reduction for load forecasting based on support vector.machines", IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies (DRPT2004), 2004.
[3]. Vapnik, V., "The nature of statistical learning", second edition, ©Springer, 1999.
[4]. Cherkassky, V. and Ma, Y., "Practical Selection of SVM.
[5]. Parameters and Noise Estimation for SVM regression. Neural networks, vol..17, PP-113-126-2004.
[6]. 6] Financial Instability Analysis using ANN and Feature Selection Technique: Application to Stock Market Price Prediction- Risul Islam Rasel, Nasrin Sultana , Nasimul Hasan, IEEE 2016
[7]. Predicting Stock Market Price Using Support Vector Regression- Phayung Meesad, IEEE 2014
[8]. [8] [Stock market index prediction using artificial neural network- Amin Hedayati Moghaddam, Moein Hedayati Moghaddam, Morteza Esfandyari, Journal of Economics,2016.
[9]. [9] Rasel, Risul Islam, Nasrin Sultana, and Phayung Meesad. "An efficient modelling approach for forecasting financial time series data using support vector regression and windowing operators." International Journal of Computational Intelligence Studies 4.2 (2015): 134-150

Smart Dustbin with Route Optimization

Neha Kapadia, Grishma Pandiyath, Palavi Patil, Sneha Singh, Dr. Rupa Mehta
Information Technology Thakur College of Engineering and Technology, Mumbai, India
Computer Department SVNIT Surat, Gujrat, India
nehap03@gmail.com, grishma097@gmail.com, palavip99@gmail.com, sneharsingh49@gmail.com, rgm@soed.svnit.ac.in

Abstract: **It is a common sight to witness garbage spilled out in and around the dust bins. The area around an improperly maintained dustbin can house disease spreading insects like mosquitoes, flies, bees and driver ants. The proposed system has smart dustbins that constantly monitor and send the level of dustbins on the cloud server where each dustbin is uniquely identified. Ultrasonic sensors are used to calculate the level of waste in the dustbin, which are transmitted via Node MCU. The dustbins are connected to the internet using Wi-Fi. Upon transmission, if the level of waste in the dustbin is more than some threshold limit, the dustbin becomes a node. Only after some nodes are formed, the route for garbage truck is formed using Shortest Path Algorithm and sent to the truck driver for garbage collection. The whole system is accessed using an Android Application. The truck drivers are notified and shown the optimized route within the application. The system is proposed for the efficient garbage collection in smart cities, with the help of smart dustbins for waste monitoring with most optimized routes based on several parameters for the truck drivers for smart garbage collection.**

Key words: Waste management, route optimization, smart dustbin, Node MCU, IOT, Cloud, Wi-Fi, Wireless, Ultrasonic sensors, Static & Dynamic Routing, Android application

I. INTRODUCTION

Today's waste management systems involve a large number of employees being appointed to attend a certain number of dustbins. This is done every day periodically. This leads to a very inefficient system in which some dumpsters will be overflowing and some dumpsters might not be even half full. This is caused by variation in population density in the city or some other random factor which makes it impossible to determine which dustbin needs immediate attention. The introduction of the sensor-based garbage dustbins is highly effective system to control the waste management in the cities. Dustbins will be embedded with ultrasonic sensors at the top, the ultrasonic sensor is used for measuring distance. This makes it possible to measure the amount of waste in the dustbin. If specific number of dustbins get full above the threshold level then an alert message will be sent to the corresponding personal. We are proposing a system for the efficient garbage collection in smart cities, with the help of smart dustbins for waste monitoring and an android app for managing the whole system by generating the most optimized routes based on several parameters for the truck drivers for smart garbage collection, and providing the GUI for other users. The smart dustbins use Node MCU and Ultrasonic sensors for monitoring the level of waste in the dustbin and uploading real-time the data on the server. Using the shortest path algorithm, will find the optimized path/route that the driver should take for emptying the dustbins considering the level of waste in dustbin.

II. BACKGROUND

Due to increase in the amount of garbage and a need to manage it, many researches have been done on waste management using Internet of things to look after the issues related to cleanliness which in turn result into bad health. Each paper has adopted a different approach and methodology. The information provided by every paper is very beneficial and helpful in taking our application to the next level.

The use of Arduino UNO brings in the need for an external Wi-Fi module which can be avoided in Node MCU along with cost saving. Node MCU is cheaper and the built-in Wi-Fi module further reduces the cost of external device.[1][2][3]

Some of the existing systems, are run entirely on solar panels. Therefore, availability of sunlight is needed most of the time. The cost of such a system is quite high since it uses solar panels and elevated assembly.[10] Piston can be attached to put additional pressure to accommodate more garbage. Elevated assembly collects the waste thrown outside and put it back in the dustbin properly.[4]

LCD module has been used to display the level of garbage in the dustbin. When dustbin is full, it sets timer. If the dustbin is not emptied within the given time, higher authorities will be informed via SMS. No route optimization is done which leads to increased fuel consumption.[5]

Algorithms like Dijkstra's shortest path algorithm, Genetic algorithm can be implemented for static and dynamic routing network. When compared, Dijkstra's algorithm takes more time in finding the optimum path as compared to Genetic algorithm.[6]

III. PROBLEM DEFINITION

To develop an android application for smart cities using smart dustbins and route optimization which will provide the following:

1. Optimized route for drivers for garbage collection
2. Real time monitoring of dustbin level

The main objective of this project is to design an effective and efficient garbage collection system for the smart cities. Other objectives of the project are: To provide dynamic routes for the truck drivers based upon traffic, road availability. To make the system fast, efficient and affordable. To connect the system over the internet for real-time monitoring and increasing efficiency. To make use of latest technologies for improving accuracy and other factors of system like cloud computing, machine learning. To minimize human efforts. To lessen the consumption of fuel, time, data, storage, etc. To spread awareness about cleanliness, hygiene and sensible use of energy and resources. To

give the citizens a clean and healthy environment to live in.

This project has huge scope in present as well as future. It can be expanded in terms of hardware, software, network and many more parameters. Time prediction can be incorporated in the system to predict dustbin filling time so as to make the system more efficient. The presence of spare digital and analog pins on the Node MCU in the smart dustbin gives the chance of adding more devices or sensors like fire sensor, temperature sensor, gas sensors, etc.

IV. RELATED WORK

Smart city can be defined as the utilization of the information and communication technologies, and recently IoT, to transform the cities from simple to smart. Such goal is achieved by exploiting different smart services of such as smart infrastructure, smart transportation, smart energy. The basic project idea is to design a smart waste detection system which would automatically notify the officials about the current status of various garbage bins in the city, would have real-time monitoring capabilities, which would be remotely controlled using IoT techniques. It can also be seen that it is mentality of some peoples that they will throw the waste into such places in spite of dustbins. The best example of this was canteen or work place like schools, colleges, government offices etc, where peoples took something and they throw them somewhere or remain them at their table where they were taking them. Most of them do same things but they didn't throw those wastes into dustbins. It is the mentality of the people's especially Indian peoples, if they will get something new which is related to technology or autonomous they feel tickle to look forward it. So, keeping all things in mind a proposed dustbin model is prepared, which is a Smart Dustbin.

Fig 1. Existing waste management system

Smart dustbin development in India: Four young men from Nashik came up with a unique idea in a bid to further their efforts to keep the country litter-free. They built smart automated waste-bins called the KRRYP Garbage ATM. The system actually has two dustbins and one LCD screen. Users will be provided with a card that has to be swiped for the dustbins to work. Once they have swiped it, the LCD screen will automatically ask the user a general knowledge question with two options as answers (for either of the bins). For instance, if the user thinks the right answer is Option B, then they will dispose their waste in the bin marked B. Depending on whether their answer is accurate, they will collect points in their smart cards, which they can later trade for coupons or cash.

An engineering student from Rampur**,** UP, has developed a dustbin with built-in sensors that would go off if garbage is thrown around the 'Smart Dustbin', compelling passersby to put it in the dustbin. The developer, has inducted a sensor instrument which is capable to catch the smell of garbage not dropped in the dustbin. It can sense the household wastes thrown on the roads. Inspired by India's Swachhata Abhiyan, the young research scholar informed media that an infra-red receiver, buzzer, transistor and a battery has been used and it took five months' time to complete the project named 'Smart Dust Bin'.

An enterprising group of students from Mysuru has come out with the solution to address the overflowing problem of garbage containers across the city. The students' team has developed a concept of smart dustbins that will enable the civic officials concerned to monitor storage level of garbage containers on their android phones through an application.

In a first in Rajasthan, three smart dustbins will be installed by the Jodhpur Municipal Corporation at different places in Jodhpur. The civic body plans to install 35 such dustbins across the city soon.

In North and South Delhi bins are installed which is fitted with sensors that send signals once the bin is full. The initiative aids in reducing hazards such as bad odor and harmful gases. Steel Authority of India Ltd (SAIL) has launched Internet of Things (IoT)-driven garbage bins across South and North Delhi Municipal Corporation area, which automatically sends a signal to the waste collection authority once the bins are full.

Kathmandu has become the first city in Nepal to install smart solar dustbins in major streets of the metropolis. Taking a major leap toward the waste management in the city, the Kathmandu Metropolitan City (KMC) has initiated the campaign to install 100 smart dustbins in the metropolis.

V. METHODLOGY

This proposed system does real time monitoring of the dustbin. The ultrasonic sensor will detect the level of garbage in the system. This data will be uploaded on the Firebase cloud using the Node MCU. The Node MCU has an inbuilt Wi-Fi module. Once, the threshold limit is reached, dustbin becomes an active node. When a specified number of dustbins become active nodes, they form a network. Then the system applies route optimization and provides optimized route for drivers to collect the garbage. This reduces overall time, fuel and cost.

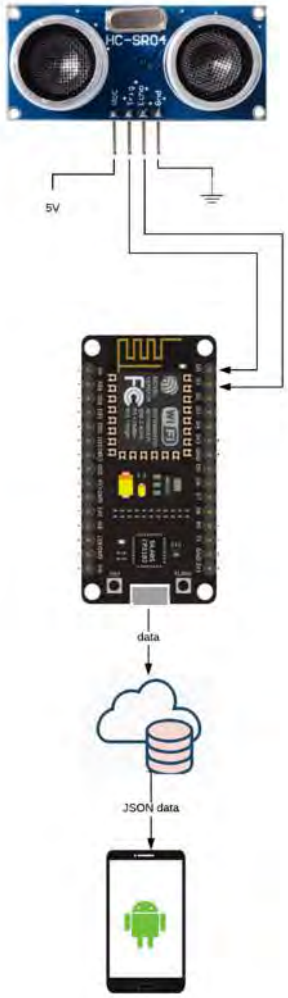


Fig 2. Block Diagram

The figure 1 explains the methodology for the implementation process.

VI. PROPOSED ARCHITECTURE

Proposed system has been implemented using following hardware:

i. Node MCU
ii. Ultrasonic sensor
iii. Jumper wires
iv. Breadboard

Fig 3. Outside the dustbin

Fig 4. Inside the dustbin

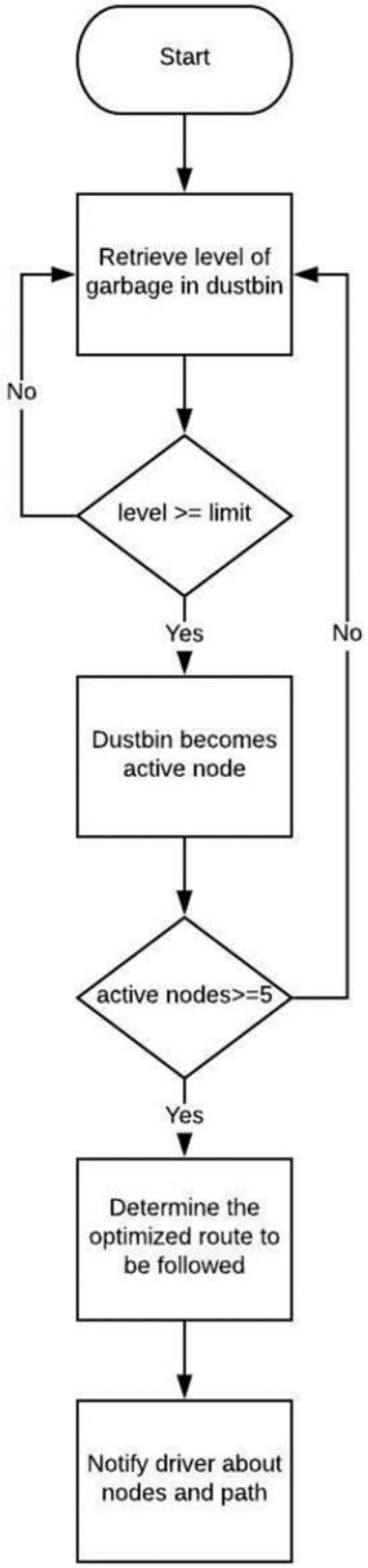


Fig 5. Flowchart

The proposed system retrieves the level of dustbin from a real-time database on periodic basis and checks whether it has reached a minimum threshold. If a certain threshold is reached, dustbin becomes an active node. When the count of active nodes is up to certain limit then optimized route is sent to the driver including all the active nodes for efficient garbage collection.

VII. FUTURE SCOPE

Fire sensor can be used to prevent any accidents

- Gas sensor can be used to detect any presence of poisonous gases in dustbins
- Time prediction: Time prediction can be incorporated in the system to predict dustbin filling time so as to make the system more efficient
- The new gen technologies and algorithms can be used in the system like artificial intelligence, future prediction and analysis, robotics, etc.

VIII. RESULTS AND DISCUSSION

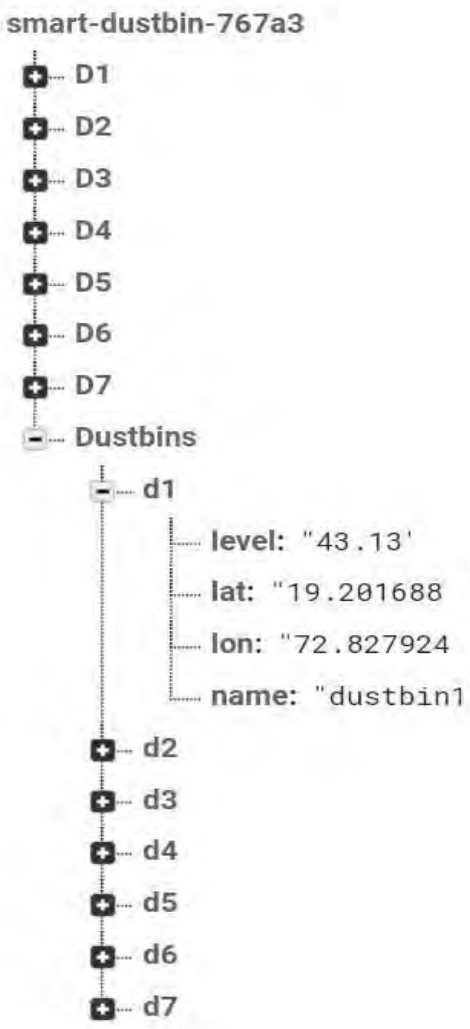


Fig 6. Firebase Cloud Database

This system uses NoSQL database for storing the data from dustbins. NoSQL databases offer storage for non-structured data and thus are flexible to query for information. The data is brought to app in JSON format which is very lightweight and compatible with various platforms along with having good amount of security. Each dustbin has level of dustbin name, latitude and longitude as child nodes.

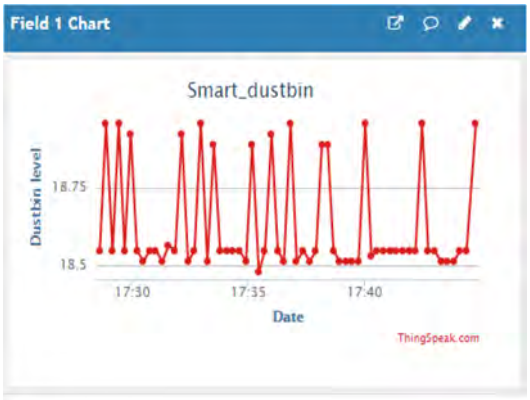


Fig 7. Graphical representation of level of dustbin

Fig 8. Locations of dustbins on the map

This image shows the locations of various dustbins in a locality. The application retrieves the coordinates of dustbins in terms of latitude and longitude and plots the coordinates on the map as markers.

The system determines the active nodes based on dustbin levels and marks them separately. A route is drawn after a certain count of active nodes is reached.
This route is displayed to the truck driver so that it can be followed for efficient garbage collection.

IX. SUMMARY

After analyzing the existing systems and their drawbacks, and researching various technical, journal and conference papers, we have finalized the hardware and software based on the shortcomings of the existing systems.

The main component used in smart dustbin is the Node MCU which is not only readily available and cheap, but also facilitates us with its built-in Wi-Fi module. Cloud technology is used to store the real-time monitored data of dustbins. Using the android application, the drivers will be provided with an optimized route for garbage collection. Also, real time monitoring of the dustbin's level will be done.

Agile methodology is used in the whole project which makes it more efficient, increases the development speed and is developer friendly as well. Moreover, the latest requirements or technologies can be easily incorporated as and when needed.

X. CONCLUSION

The increase in population, has led to tremendous degradation in the state of affairs of hygiene with respect to garbage collection system. The spill-over of waste in civic areas pollutes the neighboring areas. The proposed system has smart dustbins that constantly monitor and send the level of dustbins on the cloud server where each dustbin is uniquely identified. Ultrasonic sensors are utilized to compute the level of waste in the dustbin, which are transmitted by means of Node MCU. Upon transmission, only if the level of waste in the dustbin reaches some threshold limit, then that dustbin needs to be emptied and it becomes an active node. When a certain amount of nodes becomes active, a network will be formed.
An Android application is responsible for the overall controlling of the system and for providing the Graphical User Interface(GUI) for the users. The application retrieves data from NoSQL database in JSON format. Thus, providing lightweight data transfer. The location coordinates of dustbins will be displayed on the map and after the dustbins turn active, route will be determined and displayed to the truck drivers.
This route will be optimized considering the parameters like roads, traffic, fuel consumption, capacity of dustbins and trucks.

References

[1] Shamlee Rashinkar, Sneha Ghatole, Swati Kadapatti, Varsha Yadave, Chaitanya Jambotkar, "IoT Based Smart Trash Bins – A Step Toward Smart City", IRJET – Dec 2017
[2] Swati Sharma, Sarabjit Singh, "Smart Dustbin Management System", IJESRT – May 2018
[3] Chinmay Kolhatkar, Prachi Choudhari, Bhavesh Joshi,Dhruvin Bhuva, "Smart E-dustbin", IEEE - 2018
[4] S.Gowri, M.Divyalakshmi, S.Abinaya, C.Aarthi, "Clean India- A novel approach for waste management system using IoT", International Conference on NexGen Technologies – Jan 2018
[5] Ms. Rupa, Ms. Rajni Kumari, Ms. Nisha

Bhagchandani, Mr. Ashish Mathur, “Smart Garbage Management System Using Internet of Things (IOT) For Urban Areas”, IOSRJEN – May 2018

[6] Yagvalkya Sharma, Subhash Chandra Saini, Manisha Bhandhari, “Comparison of Dijkstra’s Shortest Path Algorithm with Genetic Algorithm for Static and Dynamic Routing Network”, International Journal of Electronics and Computer Science Engineering - 2012

[7] Nuno Garcia, Przemyslaw Lenkiewicz, Mario Freire, Paulo Monteiro, “On the Performance of Shortest Path Routing Algorithms for Modeling and Simulation of Static Source Routed Networks -- an Extension to the Dijkstra Algorithm”, ICSNC - 2007

[8] Nyan Dawood Salman, Hanan Kamal AbdulKarim, “Dynamic Shortest Path Analysis of Routing Technique via Genetic Algorithm”, Journal of Engineering and Sustainable Development – July 2018

[9] Pranay Mahajan, Avani Jain, “Implementation of municipal solid liquid based smart waste management system- zero waste using internet of things”, IJESRT – Aug 2018

[10] Abid Khan, Neju K Prince, “Advanced Dustbin System”, IJSRCSEIT - 2017

Project Management and Collaboration

Amey Tendulkar, Sneha Yadav, Shubham Rane, Mr. Sudhir Dhekane
Dept. of Information Technology
Thakur College of Engineering and Technology Mumbai,India
tendulkaramey@gmail.com, ysneha411@gmail.com, shubhamrane040@gmail.com, sudhirdhekane1987@gmail.com

Abstract—**In engineering colleges, many students from all over India work on various topics throughout their engineering course. There are chances that some topics would be similar or some component or part of topic might be similar to another topic. If students are made aware of this, then can collaborate together and reduce the duplication of efforts done behind a project and hence can focus on the other innovative part of their project. Such projects could also be incorporated in University's curriculum for smooth functioning of project by faculties in college. In current systems, colleges have to manually keep record of projects being made year by year, this manual work leaves very less scope for domain based analysis. Hence with use of Project Club will reduce that tedious work. It will provide a space where college authorities could analyze and understand what kind of projects, in regards with domain, are being developed in their colleges. This project will make use of a recommendation system to suggest similar projects based on the description uploaded by the students, this kind of idea yet remains unimplemented. We have Platforms that allow code sharing, communication, discussion forum but Project Club provides all these functionalities under one roof. Students can also seek guidance from faculty members, which will also be included as one of the entities in the proposed solution. With Faculty, Industry Personnel will also be included with the aim of providing students a direct contact with what is trending in the industry. Along with that, the presence of Industry personnel on the platform will motivate students to work as there would be chances of them getting popular. There will be a special space allotted to teams after they have collaborated on a project, to work on the project, manage their project as well as act as an interface for communication between various entities present.**

Keywords:Project,management,students,Industry,Hiring, Analytics, recommendation, Engineering, Feedback, mentor, Collaboration

I. INTRODUCTION

across India, thousands of students in various engineering colleges develop projects in various domains. There are chances that the project developed by these students might be the same

or might have some similar key features. These projects are being developed on a small scale in these colleges and students spent a lot of time developing project and including features which have already been developed by some other team in some other part of our country. This leads to duplication of efforts and waste of time students in developing an already existing project.
Lack of guidance and motivation results in students leaving the project midway. Absence

of a platform where students can discuss projects and get guidance with already working and experienced personnel hinders the development of projects. We propose a solution which acts as a platform for students to upload the description of the project. Using a recommendation system, the students will get recommended of projects similar to theirs. Further, the students can collaborate with the teams recommended to them by our recommendation system.

On a college level, many projects are being developed each semester, each year by students. These projects belong to various different domains, colleges cannot keep track of them all manually. Overall analysis becomes tedious when there's a need to list down projects in presence of some external body or committee. To tackle such cases we propose the use of analytical dashboard which will be one stop for all of analysis to be down at a college or university level.

Students can also seek guidance from faculty members, which will also be included as one of the entities in the proposed solution. With Faculty, Industry Personnel will also be included with the aim of providing students a direct contact with what is trending in the industry. Along with that, the presence of Industry personnel on the platform will motivate students to work as there would be chances of them getting popular. There will be a special space allotted to teams after they have collaborated on a project, to work on the project.

II. IMPORTANCE OF PROJECT

The main objective of the project is to provide a platform where students can collaborate and club projects, faculties can mentor students and Industry Personnel can offer opportunities to students and colleges can perform analysis on projects being made in their college/ university.
Also, as students will be recommended about similar projects there will a reduction in the development of repetitive projects.
students will be able to upload projects, get recommended, communicate and get mentored all under

one roof.
Students will be motivated to upload projects as they will be watched by Industry Personnel and might be offered of certain opportunities if their projects are liked.
Chances of students being hired by the industry personnel is also increased as they will be uploading their actual work in form of projects.

III. OBJECTIVE AND SCOPE

The following lists different types of objectives that this project aims to achieve and its scope

1. To provide a platform for students, faculties and industry personnel coexist and work on projects
2. To help in finding out similar projects
3. To help in reducing duplication of efforts in making project
4. To provide colleges an analytical dashboard which can be used to analyzed type and the number of projects being developed in colleges.
5. To help faculties to track project progress
6. Help team members manage project

IV. PROBLEM DEFINITION

Currently there is no solution to manage such a large number of projects in Engineering Colleges and to store them digitally also Colleges do not have analytical system to get insights of students , faculties and projects. Students also face problems in finding guidance from peers or faculties related to their project.
Project Club aims to solve all the above problems.

V. MODULES

A. Student

1. Student will register on Project Club.
2. Student will add his profile details like internship, certification, projects, skills
3. Every Project can have Project Repository
4. He can view his rank among other colleges students or his college students
5. Share his profile to others
6. Form team with his group members and mentor

B. Faculty

1. Faculties will register on Project Club
2. Add all Profile and professional Details
3. Accept/Reject Team request
4. Dashboard to see all team's progress at one place
5. Individual dashboard for all teams to keep detailed update
6. Receive weekly update of each team on email

C. Teams

1. Well Designed Dashboard to manage project with team members
2. View all projects details
3. send/reject connection request
4. Chance of being noticed by our industry tie ups
5. Project Repository to show project details to others
6. Share profile of team
7. Separate dashboard to discuss with connected Teams
8. Give updates to faculty from anywhere anytime

D. Analytics

1. Provides various insights of students, faculties and projects in one click.
2. Data Table of students and projects
3. Faculty – Project Allocation

E. Industry

1. We have tie ups with various companies which may offer internships or jobs by seeing student's profile and project work.
2. Companies can also collaborate with students if they are working on similar project.
3. Companies will get periodic notification from Project Club about recommended students and projects along with their profile link.

F. Django Workers

Worker is used to execute process in application which may block or take time to execute and takes user's time. Workers runs in background independently of application processes. It makes use of database to store task which are to be executed in queue with various parameters like start time, task name, end time, arguments etc.
Workers used in this application for:

a. Send email notification for user action which are important. e.g.: team formation, task allocation, updates posted etc.
b. Periodically train machine learning models on data collected in application.
c. Send weekly updates in pdf format to faculties of projects they are mentoring.
d. Send push notification on important user actions

Using Django workers improved efficiency and speed of web application and also user experience improved.

G. Recommendation System

- The use of machine learning based recommendation systems can be seen on the rise in the industry today.
- Recommendation helps users to get what they want very quickly instead of searching from list.

-We have used Content based recommendation system in this project.

- Steps followed:

a. Collected all required data which was used for recommendation.
b. Performed Tokenization on data
c. Performed stemming on output received from above step
d. Performed lemmatization on output received from above step
e. Vectorized above data using TF-IDF
f. used cosine similarity to find similar objects and save matrix received in this step
g. query similar objects based on decided threshold from above metrics

We have used Recommendation System in this project for:

a. Recommend students and projects to industry based on summary and interest of industry.
b. Recommend similar projects to team based on their abstract, keywords and domain.

H: Component Based Architecture in Django

The Component-based architecture in Django framework treats Input field, button, checkbox etc. like components, which once generated, can be used in multiple pages. This introduces reusability in the code and one does not have to rewrite code again for declaring them. If the UI has to be altered then one does not need to change it on all pages specifically, just changing at the source would replicate it throughout the application. This reduces the developer's extra work in which they would have to rewrite the same lines of code for every new page.

VI. DEPLOYMENT

Current: Currently Application is hosted on monolithic VPS cloud server with 2 cores and 4 GB ram. Database, application server, workers, machine learning all these services are running on same server instance.

Future: Depending on requests and server load, we will scale server horizontally instead of vertical.
Instead of directly making distributed system, first we will separate database to different server which will scale app to considerable percent. Depending on further load on server it is always better to separate all services like database, application, workers, machine learning, cache to different instances to serve more concurrent users.

VII. FUTURE SCOPE

This platform currently caters to the basic needs of a student to form team, collaborate with other teams and get recommendations about teams working on similar projects like themselves. Following is a list of features that can be added or already exist and can be improved upon in the future:

Video chatting provision: Using this feature, team members of the connected teams can interact with each other on a video call or conference voice call. They can also share screens of their

devices in video calls to showcase specific project related slides or dashboards

Private version control system: A private version control system will enable the connected teams to use the repository with all the features of a git system like commit, push, pull. This will help teams to keep their work private by eliminating the need to use third party git sites to save their work

Improved Machine learning algorithms: Improving the machine learning algorithm will allow the platform to provide recommendations about the projects that other team are working on, with even more accuracy. The recommendation system can be improved by training it with more amount of data that will be generated on the site by its users. The recommendation system can also be improved by putting in place a better data cleaning system that will ensure only the right data gets fed to the algorithm in its training phase

Discussion Forum: Place where anyone can post their question and can get answer to their questions from other people along with like and comment.

VIII. SUMMARY

Project Club reduces the efforts taken by students and their mentors to communicate and synchronize the development of the project. By providing utilities like a chat section, code file uploading, screenshots uploading and more, Project Club eliminates the need to juggle between third party applications. Project Club offers students a platform to connect with industry personnel and be judged by their actual work i.e. the projects. Industry personnel can either advertise their brand or company among younger academic peers, hire them based on their works and field of work. Students get exposure to the corporate world. Project Club eradicates the need of manual bookkeeping involved in keeping tracks of projects being developed by students. Analytical dashboard provides insights to colleges by using graphical representations. Project Club also provides digital storage for projects of all the projects being developed in the institution over a long period of time

RESULTS

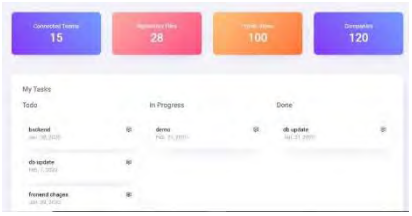


Fig1: Team Dashboard

Team Dashboard has features to manage project among team members and provide updates to faculty

Fig2: Project Management

Team Leader and Faculty can assign task to team members with deadline. Colors shows status of task. blue color: doing, red color: deadline passed, green color: done within deadline

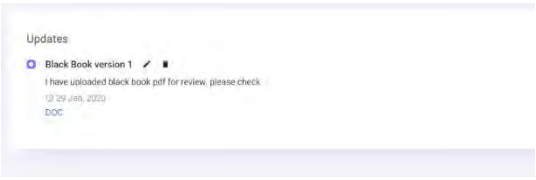


Fig3: Project Updates

Team members can post updates of project so that other members can see and comment their opinion.

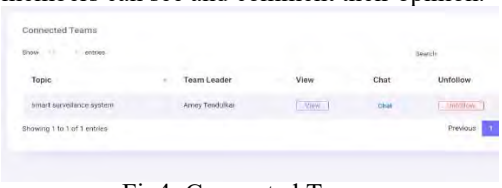


Fig4: Connected Teams

Team can discuss with other similar team to get their guidance or work related to project.

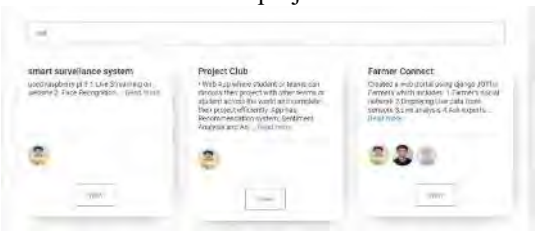


Fig5: View Projects

Team members can view projects using dynamic search and send connection request to team if they like their work

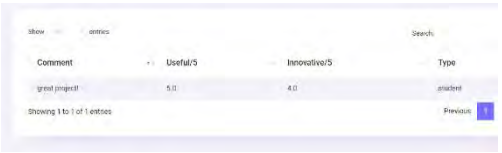


Fig6: Feedback

Receive Feedback on your project from others by sharing team profile and get to know who has given you feedback i.e.: student, faculty, industry or guest.

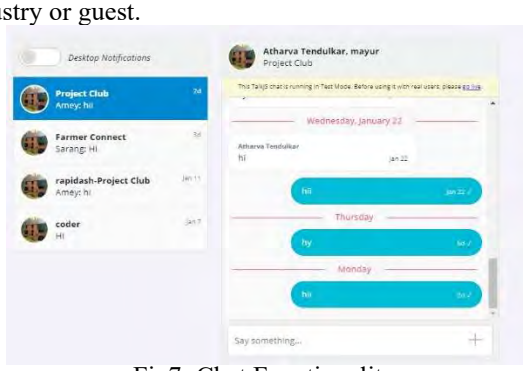


Fig7: Chat Functionality

In app chat functionality for team members to discuss project and also to discuss with connected teams.

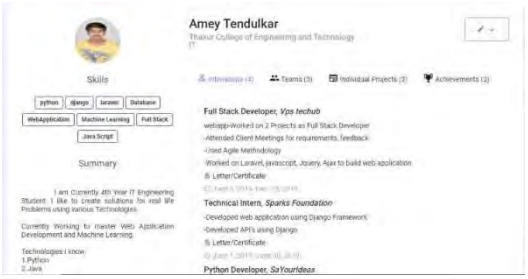


Fig8: Profile

Students can create their profile by adding their summary, skills, internships, projects and achievements and share profile with others.

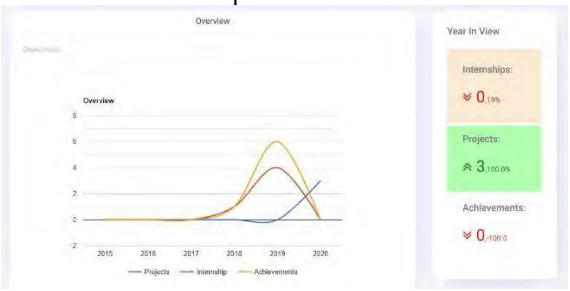


Fig9: Analytics

In depth analysis for college authorities which shows insights of projects and students of college.

Fig10: Keyword Count

This Graph shows keywords used in projects using frequency of keywords. Size of keyword proportional to frequency in projects

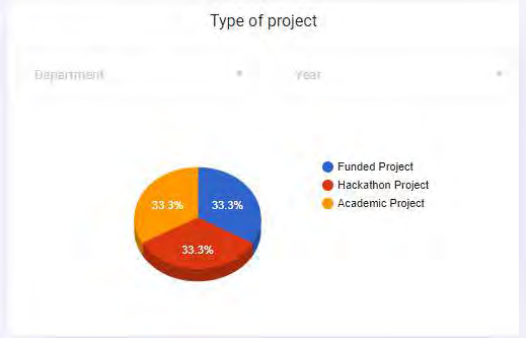


Fig11: Project Type

This Graph shows types of project going on in college Department and year wise.

IX. CONCLUSION

The use of machine learning based recommendation systems can be seen on the rise in the industry today. Recommendation systems are being used in applications for tasks ranging from suggesting which TV show to watch next to what dishes to be ordered and from which restaurant on food delivery apps.This project makes use of a similar recommendation system in helping students find other groups working on projects similar to them.

This platform will not only help students connect with other likeminded individuals, but also give them a space share codes file, communicate, get offers and mentoring all at one
place. Communication will be made easier as there will be a discussion forum present in the
project. Faculties and students can contact and discuss about a project in one place. In this way by taking advantage of all these features that the platform has to offer, students can reduce the duplication of effort build projects together, efficiently.

This platform also aims to ease the process of project management for faculties of colleges who guide students with their projects. With the student-teacher dashboard feature, students will be able to update progress they are making on projects with their respective guides seamlessly. Faculty will be able to track the progress of their groups in Realtime making the process of project management and mentoring students at college level projects hassle free.

The analytics module provides an overview of all the projects present on the platform. This helps higher authorities of colleges gain insights on the progress of projects done by the students on the platform, the domains of these projects and other trends identified by students activities on the platform. The data collection from the platform and knowledge generation done by this module will help students, faculty and industry entities to acquire a deeper understanding of the latest trends in the field of engineering by the analytical reports generated by this module.

Lastly, with the introduction of the industry entity, students will be motivated to upload more projects and increase their engagement with the platform as of this will increase their potential of getting noticed by industry personnel. This platform will benefit both the students and the industry in a mutual manner. Students will be able to showcase their talent just by being an active user the site whereas the recruiters can ease their hiring process by quickly skimming through potential employees based on skill, domains and area of interest of the individuals present on the platform.

X. ACKNOWLEDGEMENT

We sincerely thank our guide Mr. Sudhir Dhekane for his guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Rajesh Bansode the Principal, Dr.B.K. Mishra and the college management for their support

References

[1]. Prof. B Nithiya Ramesh, Aashay R Amballi, Vivekananda Mahanta, “Django the python framework”, International Journal of Computer Science and Information Technology Research, 2018
[2]. Douglas Kunda, Sipiwe Chihana, Muwanei Sinyinda, “Web Server Performance of Apache and Nginx”,IISTE, 2017
[3]. Ajay Agarwal , Minakshi Chauhan “Similarity Measures used in Recommender Systems: A Study” , International Journal of Engineering Technology Science and Research,2017
[4]. Darshita Kalyani, Dr. Devarshi Mehta, “Paper on Searching and Indexing Using Elasticsearch”, International Journal Of Engineering And Computer Science2017
[5]. R. Vidhya, G. Vadivu,Research “Document Search using Elastic Search”, Indian Journal of Science and Technology,2016
[6]. Stelios Sotiriadis, Nik Bessis, Cristiana Amza, Rajkumar Buyya, “Vertical and horizontal elasticity for dynamic virtual machine reconfiguration”, IEEE TRANSACTIONS ON SERVICES COMPUTING
[7]. 7.Django documentation: HTTPS://DOCS.DJANGOPROJECT
[8]. .COM/EN/3.0/
[9]. Sckikit-learn documentation HTTPS://SCIKIT-LEARN.ORG/STABLE/
[10]. Django-workers documentation HTTPS://PYPI.ORG/PROJECT/
[11]. DJANGO-WORKERS/
[12]. 10 https://www.analyticsvidhya.com/blog/2015/08/beginner
[13]. s-guide-learn-content-based-recommender-systems/
[14]. 11.https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243
[15]. 12.https://medium.com/beginners-guide-to-mobile-web-development/convert-django-website-to-a-progressive- web-app-3536bc4f2862

Land Registry using Blockchain

Krati Paliwal, Siddhi Poojari, Mayur Singal, Sandip Bankar
Information Technology, Thakur College of Engineering and Technology, Mumbai, India
kratip11@gmail.com, siddhipoojari7@gmail.com, mayursingal9@gmail.com, sandipbankar1@gmail.com

Abstract - **A blockchain is a growing list of records, called blocks, that are linked using cryptography. Blockchain contains a comprehensive approach in industries which requires to be protected from corruption, human error or human intervention. Land Registry is one of the use cases that involve plenty of intermediaries to place trust within the system. The present solutions in place are out of date. Land registration generally describes systems by which matters concerning ownership, possession or other rights in land will be recorded (usually with a bureau or department) to produce evidence of title, facilitate transactions and to prevent unlawful disposal. The knowledge recorded and therefore the protection provided will vary by jurisdiction .The current land registry system is rife with corruption and inefficiencies. With blockchain, we've an opportunity to mend many of those problems.**

I. INTRODUCTION

Land registration is a process by which matters concerning ownership, possession or other rights pertaining to the respective land may be recorded (usually with a office or department) to supply proof of ownership, facilitate transactions and to avoid discrepancy. Keeping track of who owns which pieces of land or property is difficult once you have hundreds or maybe thousands of years of land records and therefore the current systems in situ are outdated.

It's common to encounter discrepancies within the paperwork, including forged documents, counterfeit titles and, in some cases**,** a whole loss of all documentation. Such situations result in expensive court battles between conflicted parties. The transparent nature of blockchain can make it possible to trace how property changes hands. Blockchain's immutable, auditable and traceable features are enticing governments around the world to implement the decentralized technology within the land registry process. Using the blockchain technology it's possible to make a land registry and a history of transactions which will be easily verified at any given point. [1]

II. OBJECTIVE AND SCOPE

The objectives our project aims to achieve and its scope are:

1. Develop an accessible database to store land titles and transactions
2. To provide high capacity and throughput for millions of records
3. Data immutability that brings trust and auditability to the records
4. Quick retrieval of transaction histories

III. PROBLEM DEFINITION

Currently no solution exists to the issues faced within the existing land registry system. Thus, our system aims at building a true time record maintenance system that captures various attributes of land by using blockchain because the key technology. The system will let users to shop for and sell land securely while maintaining authenticity of the ownership. We've divided the project into 2 stages.

Stage I - will focus on shaping a base for the project. With intensive arranging, examination and structure, we'll guarantee that the important usage is smoother. Planning aims at deciding the agile methodology for arranging our land registry project. Analysis focuses on preparing an analysis of our own project. founded an itemized investigation on present enlistment and approval forms and conquering its impediments and execution benchmarks. the planning phase focuses on designing of event system application.

Stage II will concentrate on building up the undertaking and guaranteeing that the planned highlights will function as we needed them to. This stage includes developing phase where,we will send a Blockchain in which the qualifications of purchasers and venders as hashes is put away likewise in this stage we will build up an interface for our users to list our their properties and get them approved. Testing includes doing the different test framework application. The last phase is deployment where beta testing for distinguishing any further mistakes, bugs and enhancements that can be performed.

IV. TOOLS & TECHNOLOGIES

Etherium: Ethereum is a distributed public blockchain network. Ethereum specializes in running the programming code of any decentralized software. In the Ethereum blockchain miners work to earn Ether, a kind of crypto token that fuels the community. Beyond a tradeable cryptocurrency, Ether is also used by application builders to pay for transaction expenses and offerings on the

Ethereum network. There is a 2nd type of token this is used to pay miners fees for which includes transactions in their block, it is known as gasoline, and each clever contract execution requires a certain quantity of gasoline to be despatched at the side of it to trap miners to position it inside the blockchain. [5]

Metamask: MetaMask is a bridge that allows you to connect the distributed web with a normal web browser. It allows you to run Ethereum based Dapps right in your browser without implementing a full Ethereum node. It is a web browser extension that injects the Ethereum web3 API into every website's javascript context, so that Dapps can read from the blockchain.

MetaMask also lets the user create and manage their own identities, so when a Dapp wants to perform a transaction and write to the blockchain, the user gets a secure interface to review the transaction, before approving or rejecting it. Since it adds functionality to the conventional browser context, MetaMask requires the permission to read and write to any webpage.

Solidity: Solidity is an object-orientated language particularly evolved for contract writing. It is a high-stage language, which inherits traits from C++, Python, and JavaScript. The Solidity compiler compiles your source code into bytecode that runs on Ethereum Virtual Machine.Solidity is a statically-typed programming language designed for developing smart contracts that run at the EVM. With Solidity, builders are able to write programs that put into effect self-enforcing commercial enterprise logic embodied in clever contracts, leaving a non-repudiable and authoritative record of transactions. [5]

Ganache: Ganache could be a personal blockchain for Ethereum development you'll use to deploy contracts, develop your applications, and run tests. It's available as both a desktop application in addition as a command-line tool (formerly referred to as the TestRPC).

Ganache Accounts shows the accounts generated and their balances. Blocks shows each block as mined on the blockchain, together with gas used and transactions. Transactions lists all transactions run against the blockchain. Contracts lists the contracts contained in your workspace's Truffle projects. Events lists all events that are triggered since this workspace's creation. Ganache will try and decode events triggered by contracts in your Truffle project. Logs shows the logs for the server, which is beneficial for debugging. [6]

Web3: Web3 is a extensive movement and an inclusive set of protocols aiming to make the web and the net extra decentralized, verifiable, and relaxed. WEB3 is the imaginative and prescient of the serverless internet, the decentralized web. An net in which users are in control in their very own statistics, identity and future.

Remix IDE: Remix is a Solidity IDE that's used to write, compile and debug Solidity code.
Solidity is a high-level, contract-oriented programming language for writing smart contracts. It was influenced by popular languages such as C++, Python and JavaScript.

V. METHODOLOGY

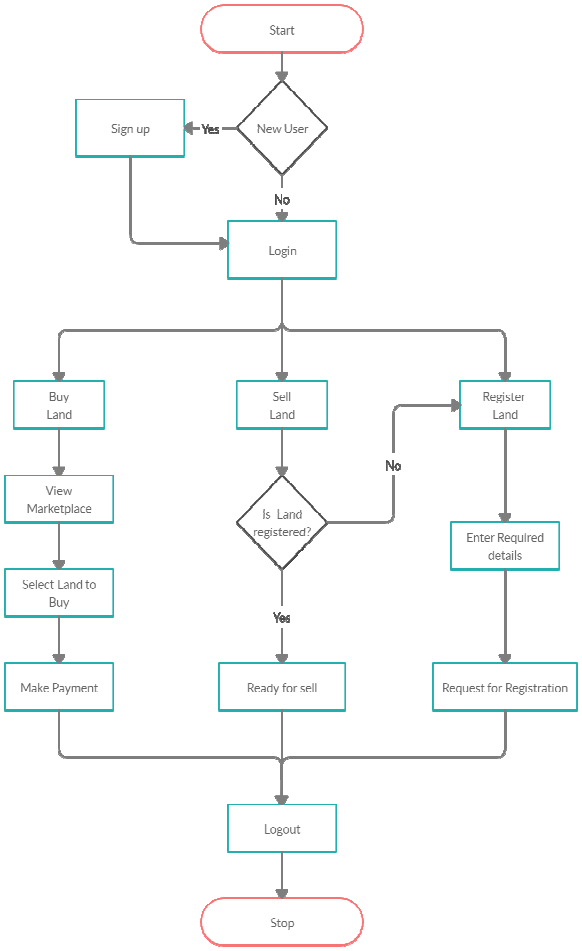


Fig 5.1: Flow chart: User [3]

User: For buying land

1. The user will register on the portal.
2. The user can view the marketplace and check the different properties listed.
3. As per his requirements the user can select the land to buy.
4. On selection of land, the payment can be made.
5. When user buys property, registration need not be requested as it will already be a part of blockchain from the seller's end.

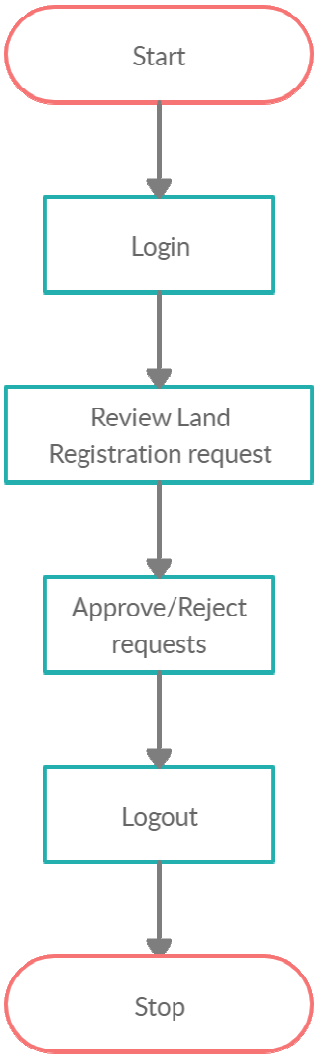


Fig 5.2: Flow chart: Registrar [3]

User: For selling land

1. The user will register on the portal.
2. For registered property, the user can list it out for sale.
3. If the property is not registered, the user will have to fill in the required details and documents and request for registration,
4. Upon completion of registration process, the property can then be listed for sale.

Registrar

1. The registrar will login into the portal.
2. He can view the registration requests on the dashboard.
3. The registrar can accept and reject requests based on the details uploaded by the requester.
4. The accepted requests will be stored as a block in the blockchain.

VI. FUTURE SCOPE

Currently the proposed system can be used to authenticate and validate property ownership and allow the user to put up properties for sale.as the verified user ownership information gets stored as a block in blockchain, it becomes secure due to security aspects of blockchain. The information will be available to different departments as per their usage. The resultant system will serve as an easy, quick and cost efficient record maintenance and retrieval system. The current project can be extended by increasing the area coverage. The proposed system can be extended to different states. The project can be further integrated with maps for better geographical identification and understanding. The system can be used as a secure and trustworthy resource to store, maintain and retrieve property information of various users over a given region.

VII. SUMMARY

The project aims at delivering a trustworthy and reliable system that will make land registration process hassle-free while maintaining security. Large amounts of records can be stored easily. The transparent nature of blockchain can make it possible to trace how property changes hands. Also, the challenges of time delays, involvement of middlemen and frauds can be overcome.

The capacities required for the venture are reasonably in the attainable range. A GUI is required, for the clients to make their records and rundown out their properties for selling or for enlisting another one. The enrollment office requires an interface for confirming the ownership. Verified data will be included to the blockchain. The GUI will be kept basic and moderate. This application will have numerous positive applications on society. It will improve the validity of the clients and give validness to proprietorship. The philosophy that we decided for our task is Agile. Light-footed enables us to be adaptable, oversee evolving necessities, deal with the regularly expanding extension just as get purchaser viewpoint.

VIII. RESULTS

Fig 8.1: Land Registry Smart Contract

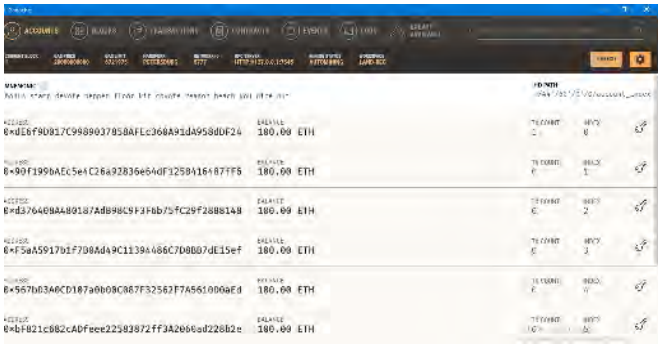


Fig 8.2: Land Registry Smart Contract [6]

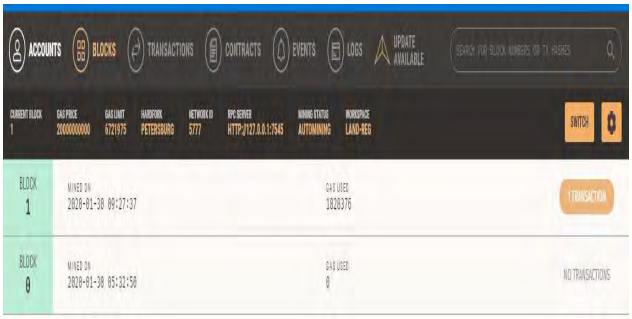


Fig 8.3: Mined Blocks

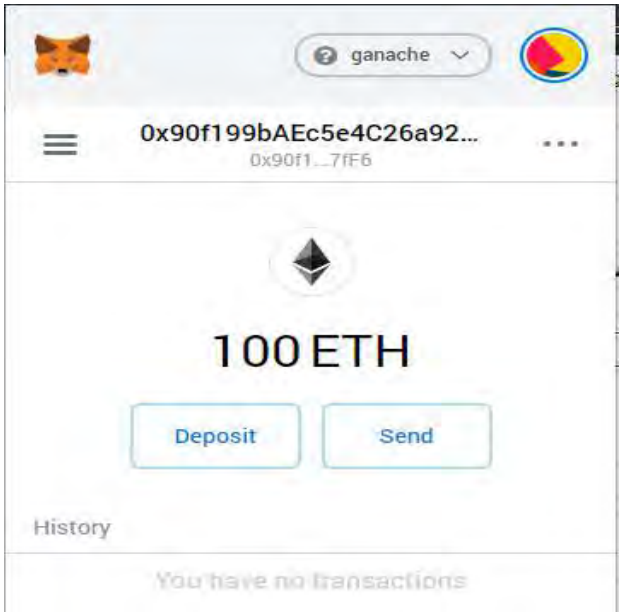


Fig 8.4: Metamask: Injected Web3 Provider

Fig 8.5: User Registration Page

IX. CONCLUSION

The Blockchain based Land registry system will facilitate integration and validation of land records over a given region. The system allows users to buy, sell and register new properties. It is easy to use and a secure way of buying and selling land. Blockchain provides the key feature of security which allows the registrar to validate the user's authenticity in a hassle freeway. As the verified user's record is entered in a block in blockchain its can surely be relied upon. The problem of discrepancies and false records is eliminated in our system. Thus, it can be guaranteed that the property listed by a user belongs to him only. This makes the system trustworthy.

References

[1] Valentina Gatteschi, Fabrizio Lambert, Claudio Demartini, Chiara Pranteda,Víctor Santamaría, To Blockchain or Not to Blockchain:That is the Question, 2018

[2] Vitalik Buterin, Ethereum: A next generation smart contract & decentralised application platform, 2014

[3] Alex Mizrahi, A blockchain based property ownership recording system, 2015

[4] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, Hawk:The blockchain model of cryptography and privacy-preserving smart contracts, 2017

[5] Maximilian Wohrer_, Uwe Zdun : Smart contracts: security patterns in the ethereum ecosystem and solidity, 2018

[6] Ganache Tool. [Online]. Available: https:// github.com/trufflesuite/ganache

Sports Analytics using Neural Networks

Samyak Tripathi, Kuldeep Trivedi, Chintan Agrawal

Abstract— **Sports Analytics includes the use of data related to sports such as players' statistics, weather conditions, information from expert scouts, etc. and build predictive models around it to make informed decisions. Analytics in the field of sports is quite necessary as historical and statistical data sets about performances of players on the field will help the managers and players evaluate their performance and make better strategies and decisions off field to improve their performance on a wholesome level. Data management tools, analytical models, information systems are all combined together for the decision-making process. Such information is primarily sought for improving the team performance. Sports analytics makes use of machine learning and neural networks to achieve its prescribed goals and working of the desired system.**

I.INTRODUCTION

The interest in sports analytics has surged, such that advances in digital technologies increases the capacity to produce data for enhancing market and leadership development in professional sports business organizations.

Sports analytics research demonstrates that creativity, framing, and robust technical analysis over intuition is behind the development of a successful sports business organization.

On the other hand, sports analytics has the potential to predict outcomes, inform management about alignment, performance improvement and business. Analytical leadership broadly refers to an analytically driven level of excellence and competitive advantage as a result of a better execution of existing analytical concepts.

Although analytics have been used widely to inform on-field performances such as player decisions, the adoption of analytics to assist other business activities in professional sports organization is often limited and anecdotal. For example, little work has been done to understand fans ' relationships to their favorite teams and athletes.

Sports analytics refers to specific use of data analytics in the industry of professional sports. Though regarded as one form of data analytics, the specific context of the sports industry refers to processes and operations involved in the sports analytics, which are different from generic business analytics or enterprise analytics.

Davenport identified three areas of activities sports analytics involve: team and player performance analytics, sports business analytics, and health and injury prevention analytics. In this sense, sports analytics entail a much more complicated and diversified process. It is for this reason that, to engage in sports analytics, professional sports teams often have to work within an ecosystem and collaborate with a range of data, software, content and services providers.

Sports Analytics is a term describing sporting operations in which a team endeavors to analyze and predict outcomes of games, analyze the market for players and buy what is undervalued and sell what is overvalued. It uses various factors for value estimation and some of them are age, performance, popularity, versatility, marketability, adaptability and a lot more.

II. PREVIOUS RESEARCH

Artificial neural networks have been studied for decades and have a significant body of literature. Even so, it has proven difficult to apply them to real world applications. The research by Purucker introduced the idea of using ANNs to predict the outcome of NFL games. This research offers an excellent introduction to using ANNs in foot- ball predictions. Purucker applied several types of supervised and unsupervised networks to predicting weeks 15 and 16 of the 1994 NFL season. Work by Kahn used 13 weeks of game data, including total yardage differential, rushing yardage differential, time of possession differential, turnover differential and home field advantage to predict weeks 14 and 15. He achieved 62.5% and 75% accuracy in these two weeks. Considering the ESPN experts in 2009 had between 61% and 67% accuracy predictions, this methodology shows promise. Loeffelholz et. Al. performed a more extensive study of neural network prediction of NBA games in. The work in this paper incorporated ideas of these past studies with several key additions including season- to-date statistics combined with previous seasons, a simple method for incorporating team statistics, a robust method of combining multiple ANNs and analysis of the performance of this methodology

III. BACKGROUND

Sports Analytics includes the use of data related to sports such as players' statistics, weather conditions,

information from expert scouts, etc. and build predictive models around it to make informed decisions. Data management tools, analytical models, information systems are all combined together for the decision-making process. Such information is primarily sought for improving the team performance.

IV. IMPORTANCE OF THE PROJECT

The sports industry uses sports analytics to increase revenue, predict outcomes, study strategies, improve player performance, calculate and evaluate player's value and a team's quality of play, prevent injury and for many more enhancements.
All this data is a great resource; however, it serves no use without people to interpret and analyses how it may be useful. Sports analysts are currently in high demand as many teams are developing entire departments just to analyses statistics. In other words, sports teams are using
Analytics for a competitive advantage.

Three specific advancements include integrating data sources to advance competition, communicate why the data is useful and create a different fan experience. The data is beneficial to many in the industry including coaches, managers, agents, scouts, marketing professionals, medical personnel and the analytics staff.
With the current available technology, sports analysts are able to take data and create insightful yet simple visualizations to communicate to other key decision makers of a team. Many teams use these models to predict the outcomes and also engage in deep analysis of oppositions performances while analysing their own as well.

V. OBJECTIVES & SCOPE

The main goal is to build a model for expected run statistics in order to better understand a team's performance and thus to generate better predictions for the future. In the process the system will test different Machine Learning techniques and algorithms in order to obtain the best possible performance.

It helps in evaluating player's value to the team and how a player may respond to various situations. It helps in analyzing player's strength and weaknesses. It will assess player's strengths and weaknesses. It will assess player's attributes such as attacking prowess, heading, defense, tackling, passing and more.

The scope of this project is:

- Predict Outcomes: Based on the historical and statistical data available, the system will be able to predict outcomes of the games based on the performances of the players and the teams.
- Coaching decisions: Coaches can use sports analytics to ascertain important and precise data sets that would help them regulate their tactics for better results on the field. Using this data, coaches and players can make more well-versed decisions that could decide wins and losses. They can also analyze the data from past matches or tournaments to frame a better game plan and remove the tactics that do not give favorable results.
- Live field data: At present, a huge amount of data is gathered manually during a game of sports matches.
- Player recruitment: Recruitment changes as young players who can be nurtured are bought or who will increase their market share with their popularity.

VI. STATISTICAL DATA

The statistics we gathered were mostly from English Premier League games throughout the course of the season. We collected a total of 46 statistics per game from this site for the EPL. We also gathered average stadium attendance, the previous season rating, and the opponent's win percentage for both home and away teams of each football game for a total of 46 statistics.

All of this was done with Perl scripts and then written to Excel files. Simple Perl modules were installed from (LWP and Spreadsheet) to work with online sources and to write to Excel files. After retrieving the raw data, we then calculated the season-to-date averages for each team. This was done using both Perl and MATLAB scripts. The results were separated by week and by the season.

The last step before using the data as input to the ANN was to average each of the expected offensive statistics with the expected defensive statistics of the opponent. So for example, if team A and team B are playing each other, instead of just using the number of goals team A is expected to score (their current season-to-date average for scoring) as in- put, we would average the number of goals team A is expected to score with the number of goals team B is expected to allow. This final number was the one that was used as input to our model. These calculations were all done with MATLAB scripts

VII. ARTIFICIAL NEURAL NETWORK

Inspired by the biological neural networks of the brain, an Artificial Neural Network (ANN) is ultimately a function, where the input is a vector of input data. This network is composed of "neurons" or processing units which are highly connected. The neurons the ANN can be taught

specific patterns by feeding the network input vectors, and targeted output vectors. Given enough data the ANN can be trained to approximate additional data vectors or find relationships between input and target vectors.
An artificial neuron can be understood as a mathematical function or object that takes input data and produces some output. Suppose that we have a n element vector as the input data, the neuron will scale each data element rj , by a weight wj . The scaled data is summed and offset by some bias b and then put through a differentiable transfer function f. The neuron output can be viewed analytically as

$$f(r1w1 + r2w2 + ... + rnwn + b). \quad (1)$$

The neuron is able to interpret the data differently depending on the weights and bias. A stronger weight for a given data point allows that data point to change the output more dramatically. The transfer function f can be linear f(x) = x, or something nonlinear. Nonlinear transfer functions can be useful for understanding nonlinear relationships. For the models used we implemented "Tan-Sigmoid" transfer functions

Now that we have introduced the basic building block of our network, we construct the feed forward three layer ANNs used in the more successful models. The first layer is constructed of 4 "hidden" neurons. In the hidden layer we implement the nonlinear transfer function "Tan Sigmoid". There exists a weight between every element of the input vector and every hidden neuron. The outputs of this layer is passed as the input vector into a hidden layer of 6 neurons with"Tan Sigmoid" transfer functions. The outputs of this layer are put through an output neuron with a linear transfer function and may be compared to the target vector and the squared error gives the error function. The error function is what we attempt to minimize with respect to the weight matrices. By optimizing the error function the ANN is being "trained".

Before we train the ANN we must first prepare both the network and the data for training. First the data is partitioned into 3 sets, the training set, the validation set, and the testing set. We use 70% for the training, 15% are used for the validation, and 15% for testing set. This is done randomly. The training data is used to create the error function. Each input and target vector creates an error function which is then averaged among all the other training data yielding the mean squared error. Iterative methods are used for optimization and the mean squared error of the validation set is plotted along with the training set. The ANN can fit the training data arbitrarily well but can easily over-fit the training data yielding higher error functions for the problem the ANN is trying to approximate. For this reason the training is stopped when the error function on the validation set is minimized. The testing data is used to test the ANN as it has not been involved in the training process. The weights and biases were updated to Levenberg-Marquardt optimization. Basic ANNs still had a good deal of variability in predictive ability. For example networks with the same structure trained on different data sets may make substantially different predictions in future seasons. This is because different ANNs partition the input data into the 3 data sets randomly. In order to create an algorithm that will give more consistent and robust results, we took an approach generally referred to as a committee of machines (CoM). In this approach many networks are trained against different random partitioning of the data set. Then based on the mean square error against the testing data, the top models are chosen to use predictively

To get the most robust approach to NFL game prediction, we used a committee of committee's approach where many committees' predictions are combined to form our final prediction. Using this approach with 500 ANNs in the training stage, the best 100 were used in each committee, then 50 such committees were used to achieve the results below. The mean was used to form each committee vote and to combine the committees' predictions.

VIII. DIFFERENT MODELS

To create an accurate model, we must either reduce the number of statistics to a manageable amount or create a large enough neural network that can find patterns in the large amount of data. We tried 5 different models for the prediction of football games. These models varied mainly in terms of the data reduction techniques implemented.

The first model we tested utilized only five different parameters. For simplicity we will refer to these statistics as on the field statistics. The statistics we got were a good starting point.

Principal component analysis (PCA) was another method we used to reduce the data being used in the model. PCA takes the 46 stats, normalizes them and creates a new lower dimensional data set made up of linear combinations of the old statistics. The new statistics are orthogonal to each other and ordered in such a way that the first principal component accounts for the maximum variability in the data. By computing this transformation, we are able to discard all but 14 principal components and keep 90% of the variance. This makes a simpler problem for the artificial neural network. This procedure also tells us that 46 independent statistics gives redundant data to differentiate games. The mathematics required was implemented using a software package in MATLAB.

Using a technique that we call linear regression combinatorial optimization we came up with a set of LRCO statistics. This method tests every possible set of statistics to identify the optimal com- bination. We started by getting the statistics from the previous season and used them to train linear regression models. The models were then tested on the following season. We repeated this procedure for every season and on every possible com- bination of statistics. The average error for each season was used to judge which statistics were best for the model. Since linear regression is many times faster than ANNs, the

computation time required to run every combination of statistics was manageable.
The results were not entirely expected. No previous research shows that 4th down conversions or attempts have a big impact on the outcome of a game. These may have been identified as useful by our models because of over-fitting or because they represent a new variable to the game. For example, scoring and time of possession both have a strong correlation of winning, but using both of them in a model might be redundant.

Another data set tested used every statistic gather. For this model we increased the size of the ANNs to 8 neurons in the first hidden layer and 4 neurons in the second hidden layer. This model was not found to be very successful.

IX. STATISTICAL ANALYSIS

One of the techniques we used in understanding our model was a derivative-based technique. In the case of ANNs, the derivatives tell us how much a change in each input will affect the output. Essentially, this will tell us which statistics are the most important. A similar methodology is used to analyze the most important factors of a deterministic model of HIV dynamics. It's a visual representation of how this technique works using ANNs.

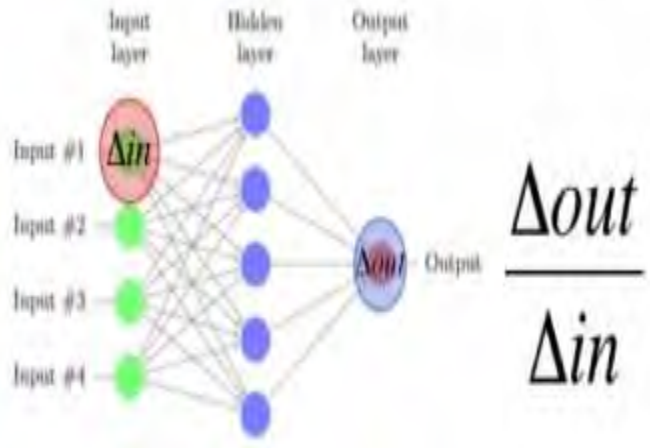


Figure 4: Visual representation of the derivative process.

Using this technique is common, however there are a few challenges using this approach on our model. Since ANNs are nonlinear functions, their derivatives will not be constant. For example, two different games will yield different derivatives. The second challenge is that our final model has 5000 ANNs. Each network will yield a different output and a different set of derivatives. To overcome this we looked at the average derivative of our model when all of the statistics are season averages. We then normalized the derivative by the input value, giving the partial derivatives the same units.

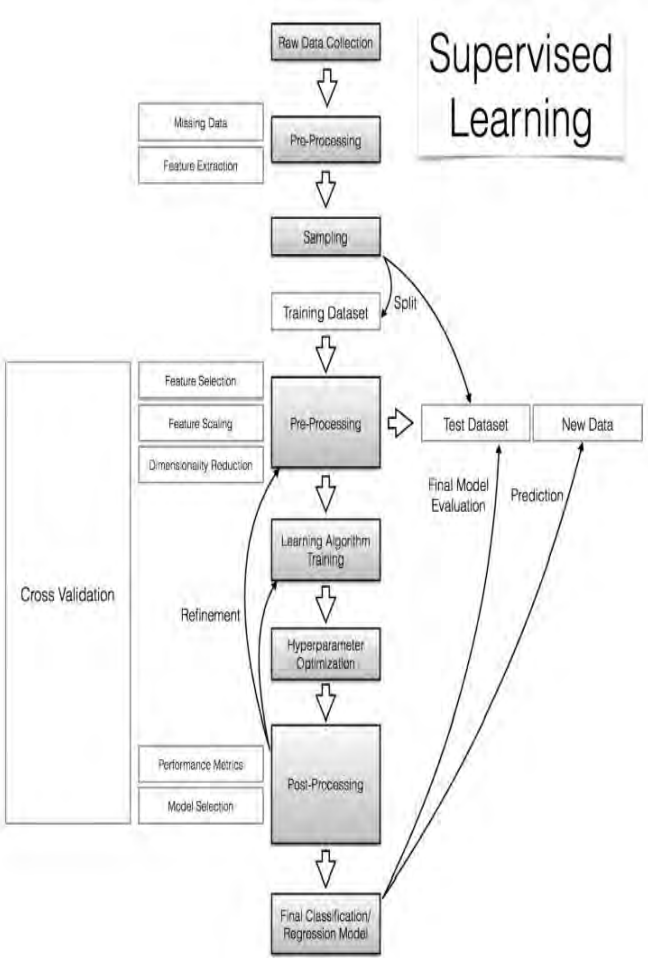
To calculate, the partial derivative was calculated for each ANN and then averaged. This was done for each statistic. Once the statistics were normalized we were able to see what percentage each statistic makes up the derivative information or the sum of the normalized derivatives.

X. PROPOSED METHODOLOGY

A. BLOCK DIAGRAM

In this paper, we propose a model to predict the outcomes of football matches in the English Premier League. We train the final dataset on various machine learning classifiers. We compare the performances of each classifier and choose the one that returns the best result. Then, we optimize the classifier that yields the best result to further enhance the model accuracy in making predictions. Our final output or target label is the Full time Match Result (FTR). This label will indicate a Home team win (H), an Away team win (A), or a Draw (D).

Architecture/Design



B. DATASET DESCRIPTION

We are going to be predicting match outcomes using data from past games for a few seasons. We obtained this data-set from football.co.uk. The data around 65 attributes per season regarding the home team, the away team, the venue, scores, to name a few. We filtered these attributes into a final list which proved to be the most influential for predicting the outcome.

C. PRE-PROCESSING

The data-set we obtained from football.co.uk consists of several attributes from each season. A lot of these features are pretty much unnecessary for making outcome predictions. Hence, our primary task is to clean the data to only retain the features or attributes that we require the most. We calculate the Scatter Matrix to observe how much one attribute affects another attribute in the dataset and their correlations. This will help us pick the most influential features that we want to use to build our new data-set.

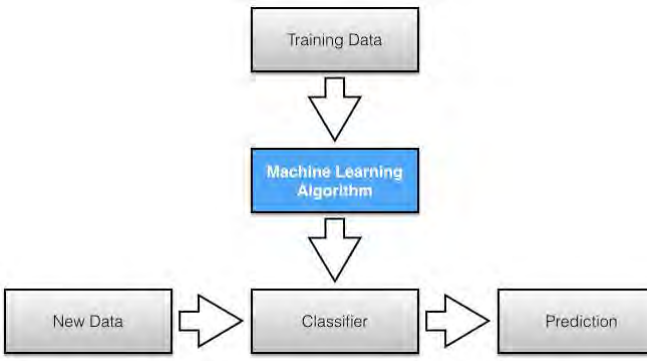
D. DATA SPLITTING:

Once we finish building our new set of crucial attributes, we split the data into training and testing data.

E. MODEL BUILDING

In this module, we apply the machine learning classifiers required for making our prediction.

F. FLOWCHART



XI. ACTUAL RESULTS

We were able to construct the required neural network using machine learning tools and successfully created a data set for 10 seasons of English Premier League using tools like rapid miner to construct accurate data sets to give us great prediction results of above 51%.

Additionally, from our findings, it was discovered that logistic regression techniques could only predict win or lost results. Further, the system could predict win, draw or loss results using DNN techniques.

The image below is our training log which shows an average of 52%-win prediction.

1	100	0.482558	1.045202	1.221429
2	200	0.523256	1.02166	1.121212
3	300	0.523256	0.991414	1.016129
4	400	0.517442	0.991745	0.925
5	500	0.517442	0.985383	0.926471
6	600	0.511628	0.978962	0.742424
7	700	0.505814	0.996009	1.5
8	800	0.517442	0.984552	0.973684
9	900	0.523256	0.996011	1.5
10	1000	0.517442	0.97818	1.225
11	1100	0.523256	0.979784	1.083333
12	1200	0.517442	0.979801	0.753256
13	1300	0.511628	0.979781	1.423077
14	1400	0.511628	0.976888	1.266667
15	1500	0.517442	0.977127	1.441176
16	1600	0.511628	0.985276	1.083333
17	1700	0.517442	0.975203	0.942308
18	1800	0.505814	0.987393	1.090909
19	1900	0.505814	0.994104	1.3
20	2000	0.511628	0.973581	1.361111
21	2100	0.505814	0.982107	1.090909
22	2200	0.505814	0.985331	1
23	2300	0.511628	0.974077	1.020833
24	2400	0.511628	0.978014	0.722222
25	2500	0.511628	0.978204	1.541667
26	2600	0.511628	0.971994	0.978966
27	2700	0.505814	0.979173	1.541667

XII. FUTURE SCOPE

Currently the project focuses on gathering a structured authentic data. As sports data and prediction is a complicated process which enlists predicting win-losses based on historical data needs to have a comprehensive and accurate database for correct predictions.

Our proposed system will need work on in enhancing the prediction rate as the current prediction rate could be better, our team will be working on increasing the prediction rate using other models and choose the one that is the most accurate one. A match making rating will be added to further boost our prediction outcomes.

XIII. CONCLUSION

Our proposed system will help us greatly in predicting match outcomes long before they are played using logical regressions and deep neural network. The AI and machine learning techniques helped us in creating results from our vast data sets that were based on data mining tools. There is a possibility that the neural network classification could be improved no matter which prediction data is used with additional training data however this additional data may be met with diminishing returns. Due to the highly volatile nature of sports, outcomes maybe difficult to predict based solely on objective data. It is possible that the additional of a human element to the prediction model would further increase its accuracy. For example, incorporating factors like team morale or saving a coach from getting fired based on the team performance in the season may add to the team's motivation to perform better and win matches that our system predicted them to lose. Our proposed system is

in no way completely accurate in predicting victory, however it indicates that an objective, statistically based system can be implemented to predict the outcomes of football matches.

REFERENCES

[1] Hucaljuk, J & Rakipović, A. (2011). Predicting football scores using machine learning techniques. 2011 Proceedings of the 34th International Convention MIPRO,, 1623-1627.

[2] Abel Hijmans. Dutch football prediction using machine learning classifiers.

[3] Darwin, P & Dra, H (2016). Predicting Football Match Results with Logistic Regression. 2016 International Conference On Advanced Informatics: Concepts, Theory And Application.

[4] Farzin, O., Parinaz, E., &Faezeh, S. M. Football result prediction with Bayesian network in Spanish league-Barcelona team. International Journal of Computer Theory and Engineering, 5(5), 2013, 812-815

[5] https://www.medcalc.org/manual/logistic_regressi on.php

[6] https://www.analyticsvidhya.com/blog/2017/09/un derstaing-support-vector-machine-example-code/

[7] https://machinelearningmastery.com/gentleintroduction-xgboost-applied-machine-learning/

[8] S.V.Manikanthan and K.srividhya "An Android based secure access control using ARM and cloud computing", Published in: Electronics and Communication Systems (ICECS), 2015 2nd International Conference on 26-27 Feb. 2015,Publisher: IEEE,DOI: 10.1109/ECS.2015.7124833.

[9] T. Padmapriya and V. Saminadan, “Inter-cell Load Balancing Technique for Multi- class Traffic in MIMO - LTE - A Networks”, International Conference on Advanced Computer Science and Information Technology , Singapore, vol.3, no.8, July 2015

Fraud Apps Detection Using Sentiment Analysis

Priyank Singh, Yash Shah, Priyanka Pundir, Shridhar Kambale
Information Technology Thakur College of Engineering and Technology Mumbai, India
singhpriyank2599@gmail.com, shahyash0505@gmail.com, priyanka.pundir09@gm mail.com, Shridharkambla1@gmail.com

Abstract—**Rank misrepresentation in the portable Application advertise alludes to extortion/misleading exercises whose lone object is to have a reason for hitting up the Applications in the prominence list.**

It turns out to be more incessant for Application designers to utilize terrible means, for example, expanding their Application deals or posting false App evaluations, to confer positioning extortion. It is vital to avoid positioning frauds there is restricted comprehension and research in this field.

Ranking extortion in the portable App business sector alludes to fake or misleading exercises which have a motivation behind knocking up the Apps in the fame list. For sure, it turns out to be more successive for App designers to utilize shady means, for example, swelling their Apps' business or posting fake App appraisals, to submit positioning extortion. While the significance of averting positioning extortion has been broadly perceived, there is restricted comprehension and examination here. To this end, in this project, we give an all-encompassing perspective of positioning misrepresentation and propose a positioning extortion recognition framework for portable Apps. In particular, we first propose to precisely find the mining so as to position misrepresentation the dynamic periods, to be specific driving sessions, of versatile Apps. This project gives a whole perspective of positioning misrepresentation and describes a Ranking fraud identification framework for mobile Apps. This work is grouped into three categories. First is web ranking spam detection, second is online review spam detection and last one is mobile app recommendation. The Web ranking spam refers to any deliberate actions which bring to selected Web pages an unjustifiable favorable relevance or importance. Review spam is designed to give unfair view of some products so as to influence the consumers' perception of the products by directly or indirectly influating or damaging the product's reputation.

Keywords: Sentiment, User Review, Fuzzy Logic System, Fake Review, Genuine Review, Fraud Application.

I. INTRODUCTION

Most of us use android Mobile these days and also uses the play store capability normally. Play store provide great number of application but unluckily few of those applications are fraud. Such applications dose damage to phone and also may be data thefts. Hence such applications must be marked, so that they will be identifiable for play

store users. So, we are proposing a web application which will process the information, comments and thee reviews of the application with natural language processing to give results in the form of graph. So, it will be easier to decide which application is fraud or not. Multiple application can be processed at a time with the web application. Also, User cannot always get correct or true reviews about the product on internet. So, we can check for more than 2 sites, for reviews of same product. Hence, we can get higher probability of getting real reviews. The recent trend in market used by the dishonest App developers for App boosting is to use fraudulent means to consciously boost their apps. At last, they also distort the chart rankings on a App store. This is usually implemented by using so-called “internet bots” or “human water armies” to raise the App downloads, ratings and reviews in a very little time. For example, Venture Beat
[1] reported that, when an App was promoted using ranking manipulation, it could be precipitated from number 1,800to the upmost 25 in Apple’s top free leader board and more than 50,000100,000 new users could be acquired within a couple of days. In actuality, such ranking fraud promotes great concerns to the mobile App industry. For example, Apple has notified of cracking down on App developers who commit ranking fraud [2] in the App store. Leading events of mobile Apps forms different leading sessions. The mobile Apps not always ranked high in the leader boards, but it usually happens in the leading sessions. So, detecting ranking fraud of mob Apps is actually the process to detect it within the leading session of the mobile Apps. Especially, this paper proposes a simple and effective algorithm to recognize the leading sessions of each mobile App based on its historical ranking records. This is one of the fraud evidence. Also, two types of fraud evidences are proposed based on Apps’ rating and review history, which gives some anomaly patterns from Apps’ historical rating and review records. In addition, we propose an unsupervised evidence aggregation method to consolidate these three types of evidences for assessing the credibility of leading sessions from mobile Apps. The main objective of this project is to develop a disease prediction system that will help the patients with finding the disease they are suffering with. It isimplemented as an application in which user answers some predefined questions and based on these questions the disease would be predicted. Along with that we can also provide some additional symptoms as inputs. The disease would be predicted using some Machine Learning algorithms. Algorithms such as Naïve Bayes, Random Forest and decision trees would help to simply the complex task of finding the disease. The margin of errors would be reduced by using such complex algorithms. It analyses the symptoms and predicts the most probable disease to the user. This application will also provide online help to the patients to get more detail about the already diagnosed disease by suggesting remedies.

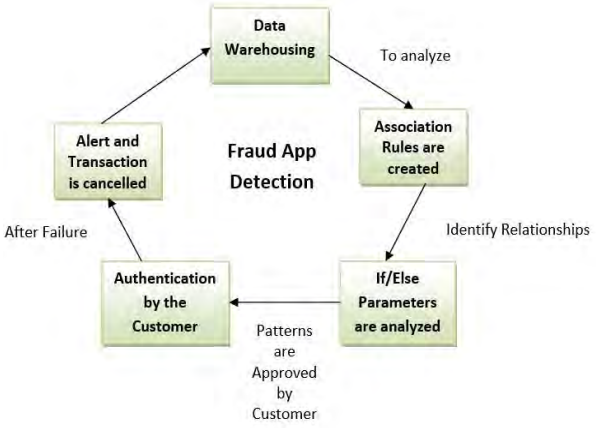


Fig 1: Fraud App Detection system

II. LITERATURE SURVEY

Liu M and others [3] Proposed that due to convenience of giant quantity of apps, apps retrieval and recommendation of apps is incredibly necessary, after that there is a need of top quality of app retrieval and recommendation there ought to be an explicit app relationship calculation for recommendation. He finished that various strategies supported app's context, which may solely find whether or not 2 apps square measure downloaded, put in meantime or offer similar functions or not.

Mohammad Aamir and others [4], explained a categorical assess of the sphere of recommender systems and Approaches for analysis of advice System to propose the
advice methodology that may any facilitate to reinforce opinion mining through recommendations.

G.Angulakshmi and others [5], discussed that the assorted tools and techniques of opinion mining, She explained that Opinion mining may be a style of tongue process that may track the mood of the folks concerning any specific product by review. Opinion mining is additionally referred to as sentiment analysis attributable to giant
volume of opinion that is made in internet resources obtainable on-line.

Vijay.B. Raut and others [6]. discussed the ways for opinion extraction, classification, and summarization. He additionally explained all approaches of summarization techniques, techniques utilized in method of opinion mining. He showed that approaches of machine learning and sentiment analysis of knowledge like motion-picture show, product, hotel etc., whereas lexicon based mostly approach is appropriate for brief text in micro-blogs, tweets, and comments knowledge on internet.

Kularbphettong. K and others [7] proposed a diamond recommendation system by mistreatment K-Means and cooperative Filtering techniques. He recommended the model system by that user mechanically so as to maximize users' satisfaction. The total system designed and developed in the form of automaton (android operative system). He divided the result by the analysis functions into a pair of parts: developing the Mobile application for diamond recommender users their evaluation and testing. The results showed that the specialists and users are happy with the system at a decent level.

Beel, J., LangerS and others [8], reviewed that of 176 publications and it showed that no agreement exists on a way to assess and compare analysis paper recommender approaches. He examined that ends up in the dissatisfactory state of affairs that despite the numerous evaluations, the individual strengths and weaknesses of the planned approaches stay
for the most part unknown.

Nikolaos Polatidis and others [9] explained the all mobile recommendation system technologies and gave the challenges long-faced the technologies. He additionally explained that mobile devices and for information retrieval various factors are crucial. He planned that development of recent Role-Based Access management (RBAC) model that improve privacy issues of permissions in apps and information accessed by user.

III. METHODOLOGY

In this section we present proposed methodology in detail. In this project, we use effective algorithm to spot the leading sessions of every App depend on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, we discover that the fraud Apps usually have completely different ranking patterns in every leading session compared with normal Apps.

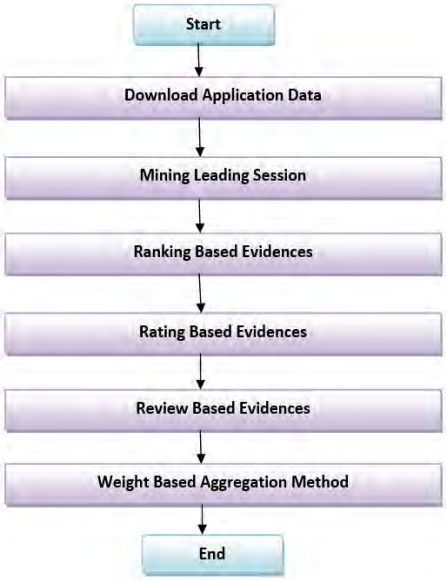


Fig 2: Fraud App Detection proposed system

1. Identifying Leading Sessions

We need to observe ranking fraud inside leading sessions of mobile Apps then propose an easy but effective algorithm to spot the leading sessions of every App depend on its historical ranking records.

After this we discover that the deceitful Apps usually have completely different ranking patterns in every leading session compared with traditional Apps.

2. Ranking based Evidences

We need to analyze the fundamental characteristics of leading events for extracting fraud evidences. By analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors in an exceedingly leading event continuously satisfy a particular ranking pattern, that consist of different ranking phases, namely, rising phase, maintaining phase and recession phase. Specifically, in every leading event, an App's ranking first will increase to a peak position within the leader board, then keeps such peak position for a time being, and finally decreases until the end of the event.

3. Rating Based Evidences

User rating is one among the utmost vital features of App promotion. An App that has higher rating could attract a lot of users to download and can even become higher on the leaderboard. Thus, rating manipulation is additionally a crucial perspective of ranking fraud.

4. Review Based Evidences

The App stores permit users to write some comments as reviews. Such reviews will project the personal perceptions and usage experiences of existing users for specific mobile Apps. Before downloading or buying a new mobile App, users read its historical reviews & based on lot of positive can download it. Thus, imposters usually post faux reviews within the leading sessions of a particular App so as to inflate the App downloads.

5. Evidence Aggregation

The study describes a ranking fraud detection process where there is some evidence considered and integrated to get an aggregated result which is most reliable in finding a fraudulent application in a mobile market [1]. Most generally the ranking fraud is happening in some particular phase of many leading events [1]. A leading event may occur due to an advertisement campaign or etc. This study can be extended to get a recommender system to enhance user experience.

IV. SENTIMENT ANALYSIS ALGORITHM

Sociologists have studied human sentiment for half a century. In the pattern of interaction between people, the meaning of vocabularies has the central role to show people's reaction to each other and also works and other actions meant to evoke a sentimental response from between vocabularies.

The increase of social media like blogs and social networks has fueled interest in sentiment analysis. So as to find the new opportunities and to manage the reputations, business folks typically read the reviews/ ratings/ recommendations and different types of online opinion. this enables to not solely realize the words that are indicative of sentiment however additionally to seek out the relationships between words in order that each word that modifies the sentiment and what the sentiment is regarding will be accurately identified.

Scaling system is employed to work out the sentiment for the words having a positive, negative and neutral sentiment. It also analyzes the following concepts to know the words and the way they relate to the conception.

There are many sentiment analysis algorithms available for developers. Implementing sentiment analysis in your apps is an easy job. There are not any servers to set up, or settings to configure. Sentiment Analysis analyzes the text of news articles, social media posts like Tweets, Facebook, and more. Social Sentiment Analysis is an algorithm that's tuned to research the sentiment of social media content, like tweets and status updates. The algorithm takes a string, and returns the sentiment rating for the "positive," "negative," and "neutral." Additionally, this algorithm provides a compound result that is an overall sentiment of the string.

For this purpose, we tend to use Classifiers is Sentiment Analysis is to see the subjective value of a text-document,

i.e. however positive or negative is that the content of a text document. Regrettably, for this purpose, these Classifiers fail to determine a similar accuracy. This can be because of the subtleties of human language, irony, context interpretation, use of slang, cultural variations and also the other ways during which opinion will be expressed. In this paper, we are using Naive Bayes classifiers.

Naive Bayes classifiers are studying the classification task from a Statistical point of view. The starting point is that the probability of a class C is given by the posterior probability P (C|D) given a training document D. Here D refers to all of the text in the entire training set. It is given by D= (d1, d2, . dn), where di is the attribute (word) of document D.

V. RESULTS

a. Outputs (Output of the project only)

Fig 3: Result

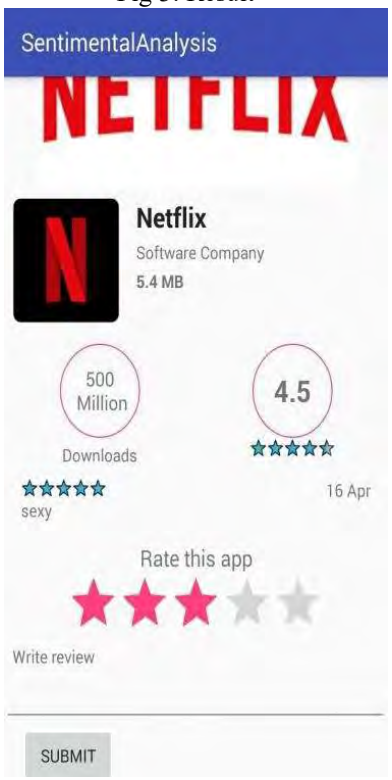


Fig 4: Result

Fig 5: Result

b. Outcomes:

The project intends to correctly analyze the input data on the basis of 14-16 parameters to correctly judge the presence of the disease. It uses complex algorithms like Naïve Bayes, Decision Tree, and Random Forest.

c. Discussion:

AI systems are now a part of a variety of specializations from stock trading to the setting of real estate prices; from fraud detection to translation of vernacular languages; from designing our daily shopping lists to forecasting which movies we might like and diagnosis of medical condition.

This is merely the beginning. Not long from today, AI will be used to advance our understanding of biological phenomenon through analysis of highly dense and huge datasets, help us invent new drugs and make treatments more thorough and personal. Self-driving vehicles can revolutionize transportation and allow new paradigms in planning of architecture. Machines will run our homes more efficiently, make businesses more productive and help predict risks to society.

Before AI systems can be deployed in healthcare applications, they need to be 'trained' through data that are generated from clinical activities, such as screening, diagnosis, treatment assignment and so on, so that they can learn similar groups of subjects, associations between subject

features and outcomes of interest. These clinical data often exist in but not limited to the form of demographics, medical notes, and electronic recordings from medical devices, physical examinations and clinical laboratory and images.

There are several techniques that can be used to correctly classify and predict when a new tuple is entered.

VI. CONCLUSION

A unique perspective of this approach is that each one the evidence is model by statistical hypothesis tests, therefore it's simple to be extended with alternative evidence from domain data to discover ranking fraud. The admin will notice the ranking fraud for mobile application. The Review or Rating or Ranking given by users is accurately calculated. Hence, a new user who needs to download an app for a few purposes will get a clear view of the present applications. Finally, we tend to validate the proposed system with in- depth experiments on real-world App information collected from the App Store. Our system will detect the ranking frauds based on three types of evidences, such as ranking based evidences, rating based evidences and review based evidences. Further, an optimization-based aggregation method combines all the three evidences to detect the fraud. This examination effectively built up an improved feeling characterization strategy for peculiarity location through web-based social networking investigation. The viability of the proposed technique is shown utilizing tweet information as a contextual investigation. The oddity estimation designs were effectively distinguished and translated through the use of the proposed technique. The contextual investigation exhibited the handiness and predominance of the technique. This exploration offers new thoughts for outlining a hearty opinion examination technique via web-based networking media information to distinguish inconsistency occasions or examples. The strategy will likewise be pertinent in cases including design changes after some

time. This ought to be considerably profitable for organizations to fortify their administration center, for political hopefuls and government pioneers to comprehend the premise of their continuous surveying comes about, and for other private associations to refine their incentives and brand guarantees to their clients.

VII. FUTURE SCOPE

In the future, we will decide to study more practical fraud evidence and analyze the latent relationship among rating, review, and rankings. Moreover, we can add more services in ranking fraud detection approach to enhance user experience. Currently the project focuses on gathering a structured authentic data. As medical data and analysis is a sensitive field, the project intends to find an assimilation of the results of various algorithms and compare the accuracy. Over the next 2 months, the team intends to build a model that best suits the input data. The model and the algorithms used are structured and developed using python as the main programming language and will use libraries like SciKit, NumPY, Django and Rasa forNLP.

The developments of the project in the future are aimed at careful consideration of the disease. This can make a prediction of the diseases by looking at a medical history of a number of patients.

Thereafter, the team intends to build a web framework to create a rich user interface.

References

[1]. Hengshu Zhu, Hui Xiong,Yong Ge, and Enhong Chen, "Discovery of Ranking Fraud for Mobile Apps"in Proc. IEEE 27th Int. Conf. Transactions on knowledge and data engineering, 2015, pp. 74-87.

[2]. A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 2011, pp. 30–38.

[3]. D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization," in Proc. 17th ACM SIGKDD Int. Conf.

"Mining personal context-aware preferences for mobile users," in Proc. IEEE 12th Int.Conf. Data Mining, 2012, pp. 1212–1217.

[17]. H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in Proc.21stACMInt. Conf. Inform. Knowl. Manage. 2012, pp. 1617–1621.

[18]. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proc.19thACMInt. Conf. Inform. Knowl.

Manage, 2010, pp. 939–948.

[19]. N. Spirin and J. Han, "Survey on web spam detection: Principles and algorithms," SIGKDD Explor. Newslett, vol. 13, no. 2, pp. 50-64, May 201.

[20]. A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf. World Wide Web, 2006, pp. 83-92.K.

Knowl. Discovery Data Mining, 2011, pp. 60–68.

[4]. A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach. Learn., 2007, pp. 616–623.

[5]. J. Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in Proc. 27th Annu. ACM Symp. Theory Comput. 1995, pp. 209–218. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072 © 2018, IRJET | Impact Factor value: 6.171 | ISO 9001:2008 Certified Journal | Page 2191

[6]. L. Azzopardi, M. Girolami, and K. V. Risjbergen, "Investigating the relationship between language model perplexity and in precision- recall measures," in Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, 2003, pp. 369–370.

[7]. N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219– 230

[8]. T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.

[9]. Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection system," in Proc. IEEE 11th Int. Conf. Data Mining, 2011, pp. 181–190.

[10]. K. Shi and K. Ali. Getjar mobile application recommendations with very sparse datasets. In Proceedings of the 18th ACMSIGKDD international conference on Knowledge discovery and data mining, KDD '12, pages 204–212, 2012.

[11]. G. Heinrich, "Parameter estimation for text analysis," Univ. Leipzig, Leipzig, Germany, Tech. Rep.,http://faculty.cs.byu.edu/~ringger/ CS601R/papers/Heinrich-GibbsLDA.pdf, 2008.

[12]. B. Yan and G. Chen. Appjoy: personalized mobile application discovery. In Proceedings of the 9th international conference on Mobile systems, applications, and services, MobiSys '11, pages 113–126, 2011.

[13]. S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In Proceedings of the 18th ACMSIGKDD international conference on Knowledge discovery and data mining, KDD '12, pages 823–831, 2012.

[14]. Z. Wu, J. Wu, J. Cao, and D. Tao. Hysad: a semi- supervised hybrid shilling attack detector for trustworthy product recommendation. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '12, pages 985–993, 2012

[15]. M. N. Volkovs and R. S. Zemel. A flexible generative model for preference aggregation. In Proceedings of the 21st international conference on World Wide Web, WWW '12, pages 479–488, 2012.

[16]. H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian,

Automated Front End Development using Deep Learning

Samdharsi Kumar, Esha Vijayvargiya, Harsh Vyas, Shruti Mathur
Information Technology
Thakur College of Engineering and Technology
Mumbai, India
samdharsi.kumar@gmail.com, eshavjy1@gmail.com, harshvyas065@gmail.com, shruti.mathur@thakureducation.org

***Abstract*—The design cycle for a web site starts with creating mock-ups for individual web pages either by hand or using graphic design and specialized mock-up creation tools. The mockup is then converted into structured HTML or similar markup code by software engineers. This process is usually repeated many more times until the desired template is created. In this study, our aim is to automate the code generation process from hand-drawn mock-ups. Hand drawn mock-ups are processed using neural networks and subsequently some deep learning methods are used to implement the proposed system.**

Keywords—convolutional neural network, deep learning, automatic code generation, HTML

I. INTRODUCTION

The importance of the Internet web sites has increased considerably due to the progress made in today's technology. Nowadays web sites reflect the face of the states, institutions, communities, people, etc. Web sites are available in almost every field, from knowledge to social work, from games to training and many more. Web sites created by companies come to the fore for financial reasons for product marketing or advertising purposes. On the other hand, official institutions aim to provide more efficient services.

At the front-end of every web site is a "web page" which is the part of the web site that interacts with the user. It is very important to serve a page that attracts the attention of the end-user, is easy to use and has enough functional features. However, developing web pages that respond efficiently to these needs involves a very burdensome process. In the preparation of web pages, graphic designers, software specialists, end-users, corporate authorities and people employed in many different areas are required to work together. Usually, the process starts with the mock-up design of the user interface by the graphic designers or mock-up artists, either on paper or a graphic editing software, in line with the needs of the institution. Software experts write code for the web pages based on these drafts. The resulting web pages may change based on feedback received by the end users. There process involves a lot of repetitive tasks. Rewriting the code for components with similar functions and page structures changing over time makes the process tedious. This reveals the need to explore more efficient solutions in web page design.

The idea of designing a web page by generating automatic code is gaining importance as a research subject. Automatic production of web pages reduces programming time, process cost and resource consumption. Thanks to the faster progressive design stages, the final web-site is produced in a shorter time.

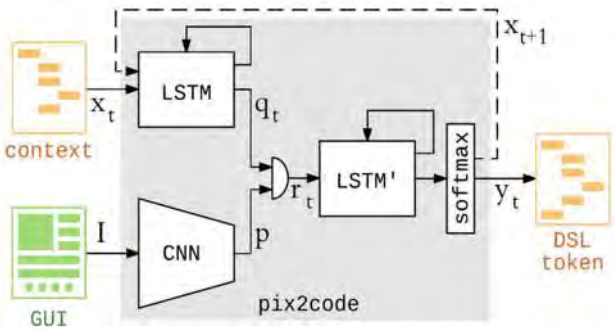
In this study, an algorithm has been developed to automatically generate the HTML code for hand-drawn mock-up of a web page. It is aimed to recognize the components created in the mock-up drawing and to encode them according to the web page hierarchy. A public dataset of hand-drawn images of web sites obtained from Microsoft AI Labs' GitHub page [13] is used to train and verify the proposed scheme. The images on the dataset are processed using a deep neural network model involving convolutional neural networks is used to train the data. Afterwards a structured HTML code is obtained. Our model achieves 96% method accuracy and 73% validation accuracy.

II. RELATED WORK

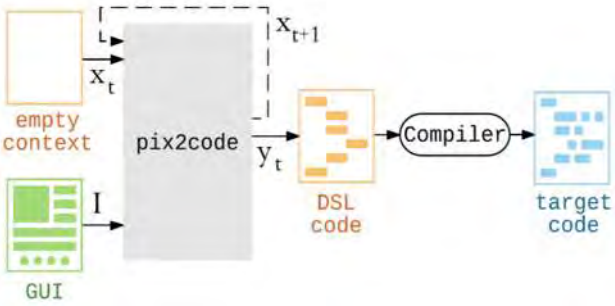
The automatic generation of programs using machine learning techniques is a relatively new field of research and program synthesis in a human-readable format have only been addressed very recently.

Pix2code is the essential work, this paper is dependent on. It is a Keras-based usage, which at its discharge likewise incorporated the dataset utilized right now. The principle distinction among pix2code and our methodology is that it is utilizing a solitary start to finish model without a pre-prepared CNN. This methodology, at its face, appears as though it would improve at the undertaking since it utilizes its own CNN prepared on interface pictures. Be that as it may, its normal exactness of 77% is lower than our methodology and Emil Wallner's. Another related work is an undertaking created by Emil Wallner, which is another Keras execution of pix2code, utilizing the equivalent dataset. It separates itself from pix2code by parting the model into an encoder and decoder, and utilizing a pre-prepared CNN as the base of the encoder model.

(a) Training



(b) Sampling



In [5], the algorithm named ReDraw takes mock-ups of mobile application screens and creates a structured XML code for it. In the first stage of their implementation computer vision techniques are used to detect individual GUI components. The second stage involve classification of the detected components according to their function, e.g. toggle-button, text-area, etc. In this stage deep convolutional neural networks are used. In the last stage, the XML code is generated by combining with the kNN

algorithm according to web programming hierarchy.

Microsoft has recently developed a system that takes hand drawn mock-ups of simple web pages and creates the structure HTML code [13]. There are no literature that explains their work, however they have published their code and dataset online. In this project we use some of the images from this dataset.

III. Dataset and Features

Our approaches require a dataset which contains a wireframe sketch and associated website code. Sourcing a quality dataset is often a challenge in many machine learning projects.

We were not aware of any dataset which contained wireframes sketches and associated website code, and therefore created our own. We considered three options to create such a dataset: (i) Finding websites and manually sketching them. (ii) Manually sketching websites and building the matching website. (iii) Finding websites and automatically sketching them. Deep learning methods require a sufficiently large dataset with hundreds of samples. Even with data augmentation techniques (augment smaller dataset into larger by modifying current samples) (i) and (ii) would require significant resources to complete.

In addition, human error and different human opinions on how exactly to create the sketches could undermine the quality of the dataset. As result we use method (iii). Our aim is to work generally on any website wireframe and therefore we required a dataset which represented a wide variety of wireframes. As such, we used real websites rather than a synthetic approach like that used by pix2code [1]. Our process involved collecting pairs of a sketched version of the website as well as the structural code of the website.

Web pages contain structure from HTML and style from CSS. We show a sample of elements with various styles. This highlights the need for a processing step to normalize web page style. We identified a number of challenges to collecting the structure from websites: (a) Wireframes have a small element set, HTML has a much larger element set, a simplified representation is required e.g. HTML has both an and tag but both represent a graphical element. (b) Wireframes are style invariant, and therefore we require only the structure. Websites have many different styles for elements. (c) Wireframes are static however; website structure can be manipulated JavaScript. (d) Wireframes only represent structure with no content. Websites contain structure and content. Content can alter the structure i.e. two web pages may have a (paragraph) element but if they had different length text the size of the paragraph would be different. As such, the structure would change. (e) HTML is sometimes invalid or poorly formatted, this would decrease our dataset quality. (f) HTML has multiple representations of the same structure. This dilutes the quality of the dataset. These issues highlight the need for careful data processing.

A. Design Mockup Features

Since we need to insert one screenshot for each word, this becomes a bottleneck when training the network. Instead of using the images, we extract the information we need to generate the markup.

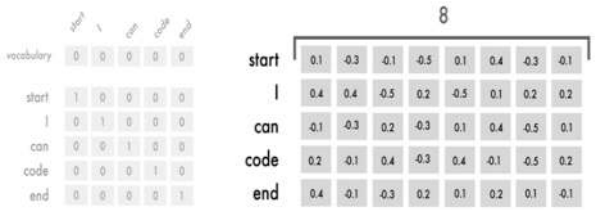
The information is encoded into image features. This is done by using an already pre-trained convolutional neural network (CNN). The model is pre-trained on ImageNet. We extract the features from the layer before the final classification.

We end up with 1536 eight by eight pixel images known as features. Although they are hard to understand by us,

a neural network can extract the objects and positions of these elements from the features.

B. Markup Features

In this version, we'll use a word embedding for the input and keep the one-hot encoding for the output.The way we structure each sentence stays the same, but how we map each token is changed. One-hot encoding treats each word as an isolated unit. Instead, we convert each word in the input data to lists of digits. These represent the relationship between the markup tags.



The dimension of this word embedding is eight but often vary between 50 - 500 depending on the size of the vocabulary.

The eight digits for each word are weights similar to a vanilla neural network. They are tuned to map how the words relate to each other. This is how we start developing markup features. Features are what the neural network develop to link the input data with the output data. For now, don't worry about what they are, we'll dig deeper into this in the next section.

IV. Proposed Method

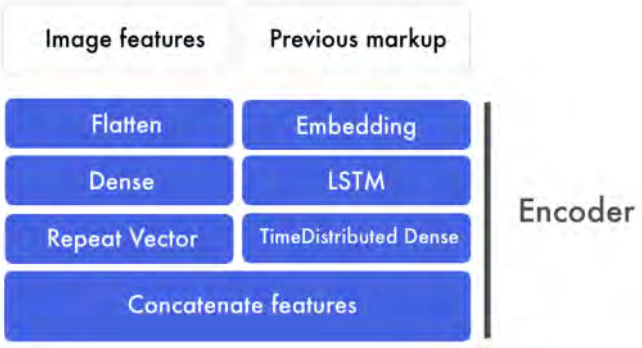
The task of generating computer code written in a given programming language from a GUI screenshot can be compared to the task of generating English textual descriptions from a scene photography. In both scenarios, we want to produce a variable-length strings of tokens from pixel values. There are two major sections. First, the encoder. This is where we create image features and previous markup features. Features are the building blocks that the network creates to connect the design mockups with the markup. At the end of the encoder, we glue the image features to each word in the previous markup. The decoder then takes the combined design and markup feature and creates a next tag feature. This feature is run through a fully connected neural network to predict the next tag.

A. The Encoder

CNNs are currently the method of choice to solve a wide range of vision problems thanks to their topology allowing them to learn rich latent representations from the images they are trained on [8, 11]. We used a CNN to perform unsupervised feature learning by mapping an input image to a learned fixed-length vector thus, acting as an encoder. We'll take the word embeddings and run them through an LSTM and return a sequence of markup features. These are run through a Time distributed dense layer - think of it as a dense layer with multiple inputs and outputs. In parallel, the image features are first flattened. Regardless of how the digits where structured they are transformed into one large list of numbers. Then we apply a dense layer on this layer to form a high-level features. These image features are then concatenated to the markup features. The input images are initially re-sized to 256 × 256 pixels (the aspect ratio is not preserved) and the pixel values are normalized before to be fed in the CNN. No further pre-processing is performed. To encode each input image to a fixed-size output vector, we exclusively used small 3 × 3 receptive fields which are convolved with stride 1 as used by Simonyan and Zisserman for VGGNet [4]. These operations are applied twice before to down-sample with max-pooling. The width of the first convolutional layer is 32, followed by a layer of width 64, and finally width 128. Two fully connected layers of size 1024 applying the rectified linear unit activation complete the vision model.

a) Markup features

Here we run the word embeddings through the LSTM



layer.

In this graphic, all the sentences are padded to reach the maximum size of three tokens.

To mix signals and find higher-level patterns we apply a TimeDistributed dense layer to the markup features. TimeDistributed dense is the same as a dense layer but with multiple inputs and outputs.

b) Image features

In parallel, we prepare the images. We take all the mini image features and transform them into one long list. The information is not changed, just reorganized.

Again, to mix signals and extract higher level notions, we apply a dense layer. Since we are only dealing with one input value, we can use a normal dense layer. To connect the image features to the markup features, we copy the image features. In this case, we have three markup features. Thus, we end up with an equal amount of image features and markup features.

c) Concatenating the image and markup features

All the sentences are padded to create three markup features. Since we have prepared the image features, we can now add one image feature for each markup feature.

After sticking one image feature to each markup feature, we end up with three image-markup features. This is the input we feed into the decoder.

B. The Decoder

Our model is trained in a supervised learning manner by feeding an image I and a contextual sequence X of T tokens xt, t ∈ {0 . . . T − 1} as inputs; and the token xT as the target label. As shown on Figure 1, a CNN-based vision model encodes the input image I into a vectorial representation p. The input token xt is encoded by an LSTM-based language model into an intermediary representation qt allowing the model to focus more on certain tokens and less on others [8]. This first language model is implemented as a stack of two LSTM layers with 128 cells each. The vision-encoded vector p and the language-encoded vector qt are concatenated into a single feature vector rt which is then fed into a second LSTM-based model decoding the representations learned by both the vision model and the language model. The decoder thus learns to model the relationship between objects present in the input GUI image and the associated tokens present in the DSL code. Our decoder is implemented as a stack of two LSTM layers with 512 cells each. This architecture can be expressed mathematically as follows:

[1] p = CNN(I)
[2] qt = LSTM(xt)
[3] rt = (q, pt)
[4] yt = sof tmax(LSTM0 (rt))
[5] xt+1 = yt

Here we use the combined image-markup features to predict the next tag.

In the below example, we use three image-markup feature pairs and output one next tag feature.

Note that the LSTM layer has the sequence set to false. Instead of returning the length of the input sequence it only predicts one feature. In our case, it's a feature for the next tag. It contains the information for the final prediction.

V. RESULTS

In this study, the purpose of generating automatic HTML code from mock-up images was successfully reached. The neural network however was trained using limited datasets and so this version will not be able to predict HTML from random websites, but it's still a great setup to explore the dynamics of the problem.
The dense layer works like a traditional feedforward neural network. It connects the 512 digits in the next tag feature with the 4 final predictions. Say we have 4 words in our vocabulary: start, hello, world, and end.

The vocabulary prediction could be [0.1, 0.1, 0.1, 0.7]. The softmax activation in the dense layer distributes a probability from 0 - 1, with the sum of all predictions equal to 1. In this case, it predicts that the 4th word is the next tag. Then you translate the one-hot encoding [0, 0, 0, 1] into the mapped value, say "end".

Example output on training the model with 'n' epochs:

10 epochs: start start start end

100 epochs: start

<HTML> <H1>Hello

World!</H1> </HTML>

<HTML> <H1>Hello

World!</H1> </HTML>

end

300 epochs: start

<HTML> <H1>Hello World!</H1> </HTML>

end

A. Output

250 epochs

450 epochs

350 epochs

VI. Conclusion

This dissertation has built upon existing research and extended this research to a novel domain of wireframe to code translation. We have presented an end to end framework which translates wireframes into websites and produces results in real time. While our work demonstrates the potential of such a system to automate the process of implementing GUIs, we only scratched the surface of what is possible.

Our model consists of relatively few parameters and was trained on a relatively small dataset. The quality of the generated code could be drastically improved by training a bigger model on significantly more data for an extended number of epochs. Implementing a now-standard attention mechanism [1, 7] could further improve the quality of the generated code. Using one-hot encoding does not provide any useful information about the relationships between the tokens since the method simply assigns an arbitrary vectorial representation to each token.

Therefore, pre-training the language model to learn vectorial representations would allow the relationships between tokens in the DSL to be inferred (i.e. learning word embeddings such as word2vec [6]) and as a result alleviate semantical error in the generated code. Furthermore, one-hot encoding does not scale to very big vocabulary and thus restrict the number of symbols that the DSL can support.

Our approaches focus on single page websites. However, they could be modified to input multi page websites. This would require our framework to be modified to accept multiple images along with information on how to link the pages together. This would allow full websites to be designed using our approaches. Finally, we focus on websites but we assert that it would primarily be an engineering challenge to use our approach described in other domains such as mobile apps or desktop applications.

References

[1] Tony Beltramelli. “pix2code: Generating Code from a Graphical User Interface Screenshot”. In: CoRR abs/1705.07962 (2017). arXiv: 1705.07962. url: http://arxiv.org/abs/1705. 07962.

[2] M. Balog, A. L. Gaunt, M. Brockschmidt, S. Nowozin, and D. Tarlow. Deepcoder: Learning to write programs. arXiv preprint arXiv:1611.01989, 2016.

[3] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

[4] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2625–2634, 2015.

[5] A. L. Gaunt, M. Brockschmidt, R. Singh, N. Kushman, P. Kohli, J. Taylor, and D. Tarlow. Terpret: A probabilistic programming language for program induction. arXiv preprint arXiv:1608.04428, 2016.

[6] F. A. Gers, J. Schmidhuber, and F. Cummins. Learning to forget: Continual prediction with lstm. Neural computation, 12(10):2451–2471, 2000.
[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.

[8] A. Graves. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850, 2013.
[9] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural computation, 9(8):1735– 1780, 1997.

[10] A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3128–3137, 2015.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[12] W. Ling, E. Grefenstette, K. M. Hermann, T. Kocisk ˇ y, A. Senior, F. Wang, and P. Blunsom. ` Latent predictor networks for code generation. arXiv preprint arXiv:1603.06744, 2016.

[13] Sketch2code. Microsoft AI Labs. [Online]. Available: https://github.com/Microsoft/ailab/tree/master/Sketch2Code/model/images

Smart Health Detection Using Data Mining

Shivam Pandey, Shivangi Singh, Surbhi Chobdar, Anil Vasoya
Information Technology Thakur College of Engineering and Technology Mumbai, India
shivampandey199888@gmail.com, shivangi2499@gmail.com, surbhichobdar98@gmail.com, Anilk.vasoya@thakureducation.org

Abstract-**Our aim is to assess the techniques of data processing in the fields of clinical and health care to develop correct choices. It conjointly offers a detailed exchange of healthful info handling ways which can improve numerous components of Clinical Predictions. It's a latest powerful technology that is of high interest in the computer world. It uses already existing info in many databases to transform it into new researches and results. From huge data sets, to extract new patterns and the knowledge related to these patterns data mining and database management. Particularly the task is to get data by the means of automatic or semi-automatic. The various parameters confined in processing of body agglomeration, forecasting, path analysis and predictive analysis. To beat the drawback of existing framework we've created good health prediction System. We have designed up a specialist framework known as good Health Prediction framework, which is utilized for improving the task of specialists. A framework checks a patient at initial level and proposes the doable diseases.**

Health monitoring systems have rapidly evolved during the past two decades and have the potential to change the way health care is currently delivered. Although smart health monitoring systems automate patient monitoring tasks and, thereby improve the patient workflow management, their efficiency in clinical settings is still debatable. This paper presents a review of smart health monitoring systems and an overview of their design and modeling. Furthermore, a critical analysis of the efficiency, clinical acceptability, strategies and recommendations on improving current health monitoring systems will be presented. The main aim is to review current state of the art monitoring systems and to perform extensive and an in-depth analysis of the findings in the area of smart health monitoring systems. In order to achieve this, over fifty different monitoring systems have been selected, categorized, classified and compared. Finally, major advances in the system design level have been discussed, current issues facing health care providers, as well as the potential challenges to health monitoring field will be identified and compared to other similar systems.

It begins with getting some information about manifestations to the patient, in the event that the framework can distinguish the fitting sickness, at that point it proposes a specialist accessible to the patient in the closest conceivable territory. In light-weight of accessible total knowledge, the framework can demonstrate the result.

I. INTRODUCTION

Sometimes we need the help of doctors immediately, but due to some reasons they are unavailable. In our project we propose a system that is user favorable to get guidance on health issues instantly through online health care system. In recent years, with reference to the various algorithms and posterior distribution the puzzles of huge data are solved highly. Meanwhile, such extracting algorithms apply to economic, social science and a few different fields.

With no additional info, classification rules are generated by the samples trained by them.

The project makes the prediction to the disease's and with the help of data mining we analysis the problem and the future solution is that we are going to provide the medicine for the disease's and the recommendation of the same.

Such associate degree outsized quantity of data cannot be processed build to form to create health predictions within the early stage and make treatment schedules to diagnose.

Our aim is to assess the techniques of data processing in the fields of clinical and health care to develop correct choices.

Data Mining is one of the most motivating area of research that is become increasingly popular in health organization. Data Mining plays an important role for uncovering new trends in healthcare organization which in turn helpful for all the parties associated with this field. This survey explores the utility of various Data Mining techniques such as classification, clustering, association, regression in health domain. In this paper, we present a brief introduction of these techniques and their advantages and disadvantages. This survey also highlights applications, challenges and future issues of Data Mining in healthcare. Recommendation regarding the suitable choice of available Data Mining technique is also discussed in this paper.

This framework improves undertaking of the specialists as well as helps the patients by giving vital help at a soonest organize conceivable. With the overall development of the web portal area the understanding of disease with the help of the user (doctor) input is the crucial aspect of till date development with the login input within the system

II. LITERATURE SURVEY

It additionally offers an in-depth exchange of healthful info handling methods which can improve varied components of Clinical Predictions.
It's a latest powerful technology that is of high interest in the computer world.
It uses already existing info in many databases to transform it into new researches and results. From huge data sets, to extract new patterns and the knowledge related to these patterns data mining uses machine learning and database management. Particularly the task is to get data by the means of automatic or semi-automatic.

1. This tells how this algorithm is used to find frequent data items and compares them with the existing algorithms.
2. Data mining techniques is used to apply on medical data which has abundant scope for improving health solutions.
3. Electronic health records and other historical medical data can prove miracles if used for a right purpose.
4. Huge amounts of complex data generated by health care sector include details about diseases, patients, diagnosis methods, electronic patient's details hospitals resources.

Our framework involves fundamental parts, for example, quiet login, enter side effects in the System, and recommend medications, proposes an adjacent specialist. The application takes the contribution of different manifestations from the patient, does the examination of the entered side effects, and gives fitting sickness expectation. Our framework enables the clients to get an examination of the indications they give for anticipating the malady they are experiencing. Some of the time the circumstance happens when you need the specialist's assistance quickly, yet they are not accessible because of some reason. Along these lines, it enables the clients to get an examination of the side effects they give for anticipating the infection they are experiencing.

III. OBJECTIVE AND SCOPE OF THE PROJECT

There square measure varied application of information mining like telecommunication business, commercial industry, biological data analysis and many more.

With the advance changes happening in the technology, especially in the field of health organization a lot of data is produced day by day. Since there is need of analysis of data and the amount of data analyzed is in large amount, so there is need of excessive knowledge regarding the technology of data mining.

For health identification E-healthcare applies data processing and telecommunication techniques. E-health was primarily used for patient knowledge analysis and wellness identification at varied levels.

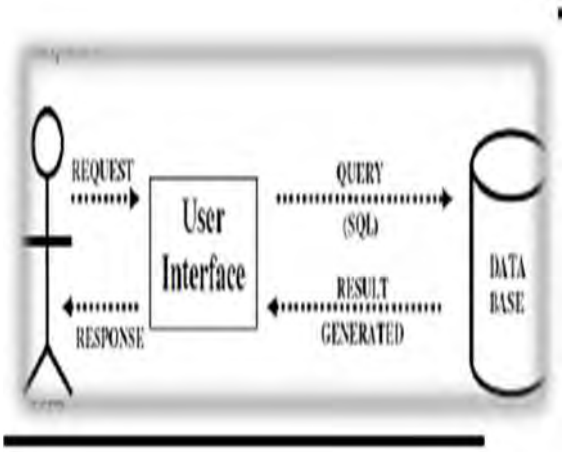
There square measure some patients World Health Organization need continuous scrutiny and would possibly would like doctor facilitate right away.

In the proposed system, hidden knowledge will be extracted from the historical data by preparing datasets by algorithm.
Predicting good health may be done solely is system responds that means.

These datasets will be compared with the incoming queries and the final report will be generated
.
Since this projected methodology can work on real historical knowledge, it will provide accurate and efficient results, which will help patients, get diagnosis instantly.
The further enhancements that can be done would be integrating this web application in an Android app.

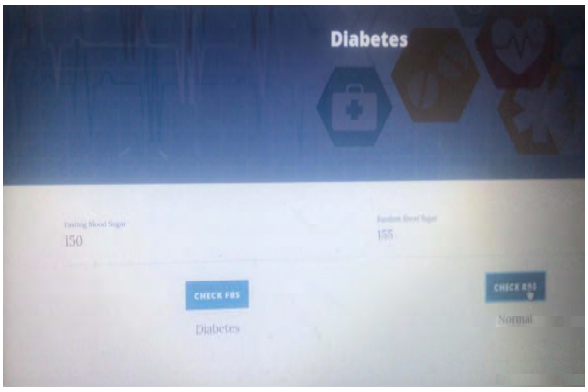
This will be obtainable to users on mobile basis and its use may be additional multiplied. Also feature like obtaining the doctor on-line on chat in order that patients will directly check with the involved doctors.
For eg the modules doing blood pressure analysis can be integrated to find how close the person associated with health issue is. This will make this web application predictable in true sense.

The further enhancements that can be done would be integrating this web application in an Android app. This will be available to users on mobile basis and its use can be further increased. Also feature like getting the doctor online on chat so that patients can directly talk to the concerned doctors. The modules doing different disease analysis can be integrated to find how close the person associated with disease is. This will make this web application predictable in true sense

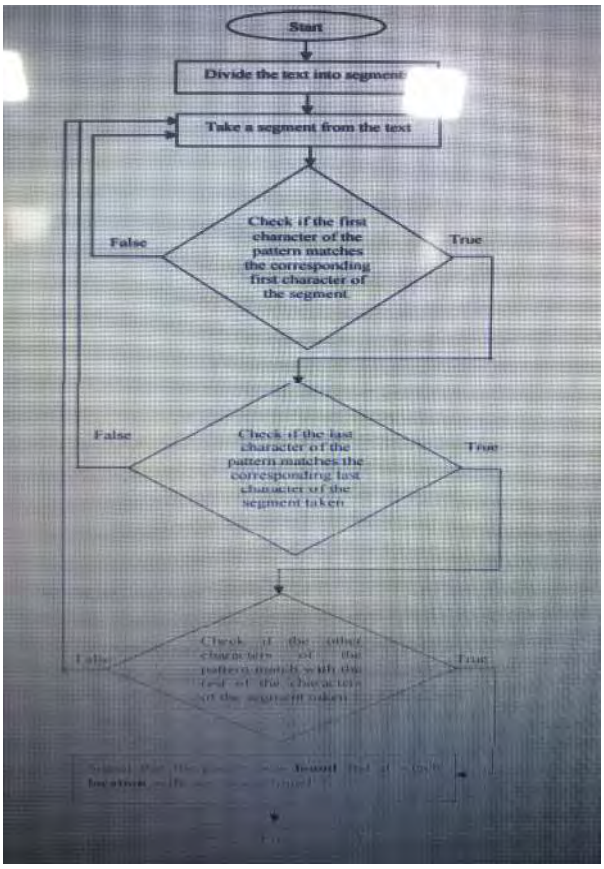


IV. PROPOSED METHODOLOGY

A. Block Diagram:



We have created an online consultation project which will help end user to get instant guidance for their health. In this system, we have imposed various symptoms and diseases which are related to them.
In the above figure the project represent as the details of patients is manually added to the dataset and with the
help



of data mining we analyses the problem and the amount of symptoms.

Here, there will be doctor's panel, Admin panel and Patient panel.
Admin can access all the information and will let patient know their requirements.

Patient will login into the system it will register and then can search various medicines and provide symptoms. Then all the disease's related to those symptoms will appear, this will be the guidance for healthcare. Doctors can view who searched for them and what they searched about.
Doctors can add various doctors and they can recommend various medicines. There will be online quesnaire andit will help the patient to understand and choose between the best.

For accuracy we are using the intelligent data mining techniques. This will help us in the accurate data sets and most accurate probability.

Health is one of the most important assets of our life which directly reflects in any form of progress or development. In today's hustle and bustle of life, most of the people neglect this asset which may be due to lack of time and complexity in the vast data available over the Internet. Data mining has introduced various techniques which would overcome this problem and assist us to emphasize on both health and work simultaneously. In present era, Data Mining is

becoming popular in healthcare field because there is a need of efficient analytical methodology for detecting unknown and valuable information in health data.

V. ACTUAL RESULTS

A. Outputs (Output of the project only)

Successfully implemented the consultation project and helped the patients to consult the doctors online and get the proper guidance.
In our project we propose a system that is user favorable to get guidance on health issues instantly through online health care system. In recent years, with reference to the various algorithms and posterior distribution the puzzles of huge data are solved highly. Meanwhile, such extracting algorithms apply to economic, social science and a few different fields.
With no additional info, classification rules are generated by them.

B. Outcomes:

Such associate degree outsized quantity of data cannot be processed build to form to create health predictions within the early stage and make treatment schedules to diagnose.
Our aim is to assess the techniques of data processing in the fields of clinical and health care to develop correct choices.
It additionally offers an in-depth exchange of healthful info handling methods which can improve varied components of Clinical Predictions.
It's a latest powerful technology that is of high interest in the computer world.
It uses already existing info in many databases to transform it into new researches and results. From huge data sets, to extract new patterns and the knowledge related to these patterns data mining uses machine learning and database management. Particularly the task is to get data by the means of automatic or semi-automatic.

C. Discussion of the results

In the proposed system, hidden knowledge will be extracted from the historical data by preparing datasets by applying algorithm.
Predicting sensible health is done solely is system responds that method.
These datasets will be compared with the incoming queries and the final report will be generated using Mining.
Since this projected methodology can work on real historical information, it'll offer correct and economical results, which can facilitate patients get identification instantly.
The application takes the contribution of different manifestations from the patient, does the examination of the entered side effects, and gives fitting sickness expectation. Our framework enables the clients to get an examination of the indications they give for anticipating the malady they are experiencing. Some of the time the circumstance happens when you need the specialist's assistance quickly, yet they are not accessible because of some reason. Along these lines, it enables the clients to get an examination of the side effects they give for anticipating the infection they are experiencing.

VI. FUTURE SCOPE

In the future, with the advent of remote monitoring devices and wireless communication systems, information/communications systems have the potential to support continuous monitoring of a patient's health status at home, rapid diagnosis by clinicians, and timely, effective therapeutic interventions in the home by the patient or a family member, with guidance by health professionals. Furthermore, by capturing process and system performance data for systems analysis, control and design, information/communications technologies can facilitate the use of systems-engineering tools by patient care teams, provider organizations, and environmental actors at all levels of the health care delivery system. Will add various chat systems and mapping will provide diet plans and locate various conference chatting.

Data mining can be helpful in the field of restorative space. Anyway protection, security and unfit to sign into the record are the huge issues on the off chance that they are not tended to and settled appropriately. It portrays the proposition of a crossover information mining model to separate arrangement learning for the guide of different maladies in the clinical choice framework and presents a structure of the apparatus different devices utilized for investigation. Now and again the circumstance happens when you need the specialist's assistance promptly, however they are not accessible because of some reason.

One of the most significant challenges of the data mining in healthcare is to obtain the quality and relevant medical data. It is difficult to acquire the precise and complete healthcare data. Health data is complex and heterogeneous in nature because it is collected from various sources such as from the medical reports of laboratory, from the discussion with patient or from the review of physician. For healthcare provider, it is essential to maintain the quality of data because this data is useful to provide cost effective healthcare treatments to the patients. Health Care Financing Administration maintains the minimum data set (MDS) which is recorded by all hospitals

For effective utilization of data mining in health organizations there is a need of enhance and secure health data sharing among different parties. Some propriety limitations such as contractual relationships among researcher and health care organization are mandatory to overcome the security issues. There is also a need of standardized approach for constructing the data warehouse. In recent years due to enhancement

of internet facility a huge dataset (text and non-text form) are also available on website. So, there is also an essential need of effective data mining techniques for analyzing this data to uncover hidden information.

VII. CONCLUSION

The system would drastically reduce the human effort, reduce the cost and time constraint in terms of human resources and expertise, and increase the diagnostic accuracy. The prediction of diseases using Data Mining applications is a challenging and risky task as the data found are noisy, irrelevant and massive too.

In this situation, data mining tools come in handy in exploring of knowledge of the medical data and it is quite interesting.

This system will also guide the users of how to remain healthy and fit using tips provided here. The further enhancements that can be done would be integrating this web application in an Android app.

This will be out there to users on mobile basis and its use may be more inflated.

Also feature like obtaining the doctor on-line on chat in order that patients will directly sit down with the involved doctors

we can also conclude that there are no single data mining techniques which give consistent results for all types of healthcare data. The performance of data mining techniques depends on the type of dataset that we have taken for doing experiment. So, we can use hybrid or integrated Data Mining technique such as fusion of different classifiers, fusion of clustering with classification or association with clustering or classification etc. for achieving better performance.

It is difficult to acquire the precise and complete healthcare data. Health data is complex and heterogeneous in nature because it is collected from various sources such as from the medical reports of laboratory, from the discussion with patient or from the review of physician. For healthcare provider, it is essential to maintain the quality of data because this data is useful to provide cost effective healthcare treatments to the patients. Health Care Financing Administration maintains the minimum data set (MDS) which is recorded by all hospitals

The modules doing different disease analysis can be integrated to find how close the person associated with disease is. This will make this web application predictable in true sense.

REFERENCES

[1]. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 10, OCTOBER 2013. “Data Mining Applications in Healthcare Sector: A Study”.

[2]. Indian Institute of Information Technology, Allahabad, India. “A survey on Data Mining approaches for Healthcare”.

[3]. Assistant Professor, Department of Computer Applications, Anna University (BIT Campus), Trichy, India P.G. Student, Department of Computer Applications, Anna University (BIT Campus), Trichy, India. “Association Technique on Prediction of Chronic Diseases Using Apriori Priori Algorithm”.

[4]. Department of Computer Engineering, RamraoAdik Institute of Technology Nerul, Navi Mumbai. “Estimation based on Data Mining Approach for Health Analysis”.

[5]. School of Information Technology and Engineering, VIT UNIVERSITY, Vellore632014, Tamil Nadu, India. “APRIORI algorithm based medical data mining for frequent disease identification”.

[6]. Chairman of CIT Faculty of Computing and Information Technology King Abdulaziz University, Jeddah Kingdom of Saudi Arabia. “Efficient Implementation of FP Growth Algorithm-Data Mining on Medical Data”

[7] S. Soni and O. P. Vyas, “Using Associative Classifiers for Predictive Analysis in Health Care Data Mining”

[8] L. Duan, W. N. Street & E. Xu, Healthcare information systems: data mining methods in the creation of a clinical recommender system, Enterprise Information System.

[9] O. Mary K., Mat, “Application of Data Mining Techniques to Healthcare Data”, Infection Control and
Hospital Epidemiology.

[10] J. I. Huircánet, "ZigBee-based wireless sensor Network localization for cattle monitoring in grazing fields", *Compute. Electron. Agricult.*, vol. 74, no. 2, pp. 258-264, 2010.

[11] R. N. Handcock et al., "Monitoring animal behaviour and Environmental interactions using wireless sensor Networks GPS collars and satellite remote sensing", *Sensors*, vol. 9, no. 5, pp. 3586-3603, 2009.

[12]. N. Kularatna, B. H. Sudantha, "An environmental air Pollution monitoring system based on the IEEE 1451 standard for low cost requirements", *IEEE Sensors J.*, vol. 8, no. 4, pp. 415-422, Apr. 2008.

[13] D. Wobschall, "Networked sensor monitoring using the universal IEEE1451 standard", *IEEE Instrum. Meas. Mag.*, vol. 11, no. 2, pp. 18-22.

[14] M. Sveda, R. Vrba, "Integrated smart sensor Networking framework for sensor-based appliances", *IEEE Sensors J.*, vol. 3, no. 5, pp. 579-586, Oct. 2003.

[15] A. Kumar, G. P. Hancke, "Energy efficient environment monitoring system based on the IEEE 802.15.4 standard for low cost requirements", *IEEE Sensors J.*, vol. 14, no. 8, pp. 2557-2566, 2014.

[16] B. C. Baker, "AN685-thermistor in single supply Temperature sensing circuits", 1999.

[17] M. Janzekovic, J. Prisenk, B. Mursec, P. Vindis, D. Stajnko, F. Cus, "The art equipment for measuring the horse's heart rate", *Achievements Mater. Manuf. Eng.*, vol. 41, no. 1-2, pp. 180-186, Jul./Aug 2010.

[18] T. Lindström, I. I. Redbo, "Effect of feeding duration an drumen fill on behaviour in dairy cows", *Appl. Animal Behaviour Sci.*, vol. 70, no. 2, pp. 83-97, 2000.

[19] E. Lindgren, "Validation of rumination measurement Equipment and the role of rumination in dairy cow Time budgets", 2009.

[20] I. Korhonen, J. Pärkkä, M. Van Gils, "Health monitoring in the home of the future", *IEEE Eng. Med. Biol. Mag.*, vol. 22, no. 3, pp. 66-73, May/Jun. 2003.

IPL Win – Loss and Score Predictor

Mr. Sanket Singh, Mr. Vivek Vikram Singh, Mr. Ankit Sharma, Mr. Vijaykumar Yele

TCET-IT, Mumbai, India.
Email id: sanketsingh98.5@gmail.com, singhvinod.vivek@gmail.com, ankitsharma2326@gmail.com, vijaypyele@gmail.com

***Abstract*- — The goal of this project is to use variety of algorithms on a comprehensive dataset of cricket match results to predict the team agnostic match outcomes. The target competition was The Indian Premier League (IPL) which is a twenty over format cricket league. We have extracted the results of every IPL match from the years 2008 to 2018. This data allowed us to calculate arbitrary batsman and bowler statistics before every game. The models were optimized with individual player features and then both team and player features. The models were trained based on the data of 11 IPL seasons from 2008 to 2018. The optimal model was a simple prediction method which can predict the outcome of a match based on the opponents and the players in particular team. Although it is difficult to take into account all features that influence the results of the matches, an attempt to find the most significant features is made and various classifiers are tested to solve the problem.**

Overview:

I. Introduction:

As a sport cricket is played globally across 106 member states of the International Cricket Council (ICC), with an estimated 1.5 billion fans worldwide (ICC, 2012-2013).However, much of the global finance and interest is focused upon the 10 full ICC member nations and more specifically upon 'the big three' of England, Australia and India .Cricket is the India's most popular sport. It is played, watched, and enjoyed by billions of people country wide. And in India particularly, the sport has more in common with religion than entertainment.

The IPL is the most-attended cricket league in the world and ranks sixth among all sports leagues. In 2010, the IPL became the first sporting event in the world to be broadcast live on YouTube. The brand value of IPL in 2017 was US$5.3 billion, according to Duff & Phelps. According to BCCI, the 2015 IPL season contributed ₹11.5 billion (US$182 million) to the GDP of the Indian Economy. So, one can easily understand that Cricket is a multi-billion-dollar market. Although Cricket is a game of uncertainties there is a strong incentive for models that can predict the outcomes of matches and beat the odds provided by bookers. The aim of this study is to investigate to what degree it is possible to predict the outcome of cricket matches.

II. Background

The main objective of this project is to develop an IPL win-loss and score predictor that will help the Management to find the best combination of player. The goal of this project is to use variety of algorithms on a comprehensive dataset of cricket match results to predict the team agnostic match outcomes. The target competition is The Indian Premier League (IPL) which is a twenty over format cricket league. We will be extracting the results of every IPL match from the years 2008 to 2018. This data will allow us to calculate arbitrary batsman and bowler statistics before every game. The models will be optimized with individual player features and then both team and player features. The models will be trained based on the data of 11 IPL seasons from 2008 to 2018. The optimal model was a simple prediction method which can predict the outcome of a match based on the opponents and the players in particular team. Although it is difficult to take into account all features that influence the results of the matches, an attempt to find the most significant features is made and various classifiers are tested to solve the problem.

III. Importance of the project

Project will able to predict the probability of a particular team winning a certain T-20 match of IPL based on the input by the user. The project will help to decide how much a given team has the chance of winning at a particular venue considering the given conditions. It will help to form a better team which is most suited to face the opponent team and increase the overall chances of winning. Using neural networks and machine learning our project aims to deliver an application to predict the success of a team in an IPL match.

Objectives and Scope of the project:

The aim of the project is to predict the probability of a particular team winning a certain T-20 match of IPL based on the input by the user. The input will take consideration of the following factors: teams playing the match, toss winning team, toss decision and the venue where the match is being played.

The input will go through several algorithms in order to predict the outcome of the match i.e. team winning the match. It will also try to predict the probable score

range for the winning team using neural networks based on the attributes passed to it. The project will help to decide how much a given team has the chance of winning at a particular venue considering the given conditions. It will help to form a better team which is most suited to face the opponent team and increase the overall chances of winning.

Other scopes of this project are:

Using neural networks and machine learning our project aims to deliver an application to predict the success of a team in an IPL match.
To predict the success of the team we are taking the statistics of every innings of batsman as well as the bowler through every IPL season from 2008-2018. Also, the team statistics at a particular venue are considered.
The prediction will be based on the outcomes to all the IPL matches in the past 11 years.
Since neural networks work on experience, we will be providing more than 600 different matches to our network, each with complete ball by ball data. This will help to gain accurate results. The number of matches won with a particular team combination, the average score of the team as well as the progress of the team through the season will be the deciding parameters.

Summary:

The project is to develop a "IPL win-loss and score predictor" this will help the Management to find the best combination of player. This project will use variety of algorithms on a comprehensive dataset of cricket match results to predict the team agnostic match outcomes. The target competition is The Indian Premier League (IPL) which is a twenty over format cricket league. We will be extracting the results of every IPL match from the years 2008 to 2018. This data will allow us to calculate arbitrary batsman and bowler statistics before every game. The input will go through several algorithms in order to predict the outcome of the match i.e. Team winning the match. It will also try to predict the probable score range for the winning team using neural networks based on the attributes passed to it.
The project will help to decide how much a given team has the chance of winning at a particular venue considering the given conditions. It will help to form a better team which is most suited to face the opponent team and increase the overall chances of winning.
Using neural networks and machine learning our project aims to deliver an application to predict the success of a team in an IPL match. To predict the success of the team we are taking the statistics of every innings of batsman as well as the bowler through every IPL season from 2008-2018. Also, the team statistics at a particular venue are considered.
The prediction will be based on the outcomes to all the IPL matches in the past 11 years. Since neural networks work on experience, we will be providing them various different matches to our network, each with complete ball by ball data. This will help to gain accurate results.

IV. Design and Implementation:

Figure No 1: Proposed Methodology

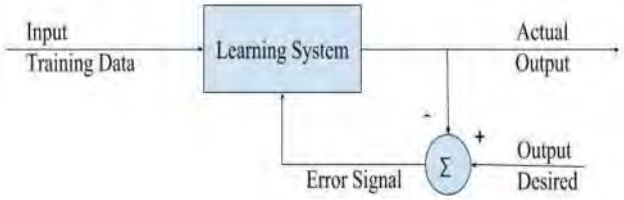


Figure No 2: Block diagram of Supervised Learning

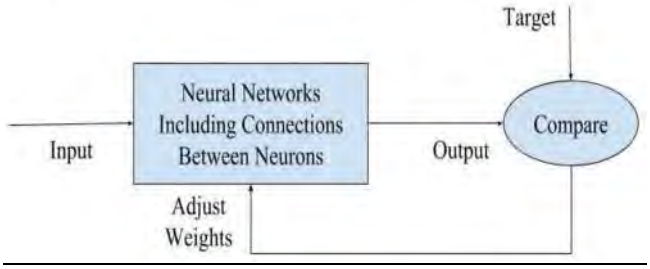
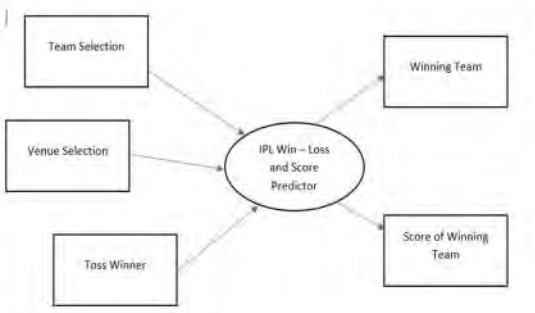


Figure No 3: Block diagram of Unsupervised Learning



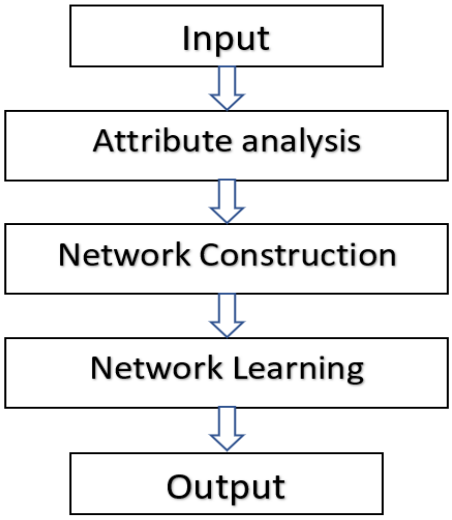
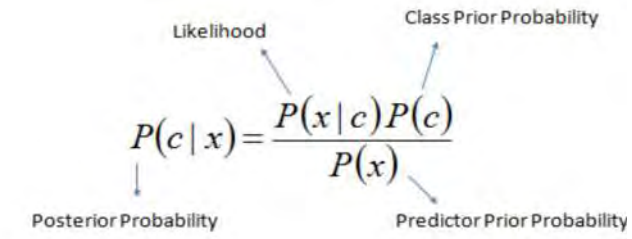


Figure No 4: DFD

V. Algorithms

It is a classification technique based on Bayes Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c$$

Figure No. 5: Naïve Bayes Algorithm

Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class.

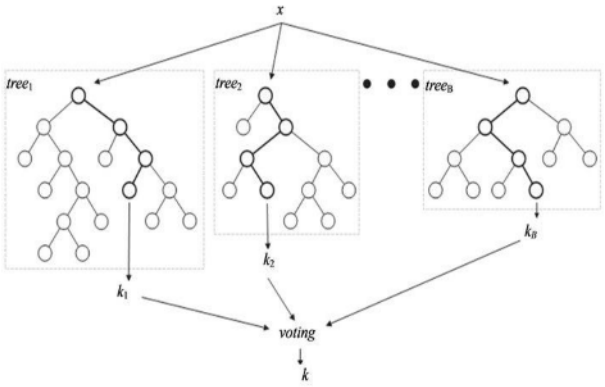


Figure No. 6: Random Forest Tree

A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are more commonly used in classification problems and as such, this is what we will focus on in this post. SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes, as shown in the image below.

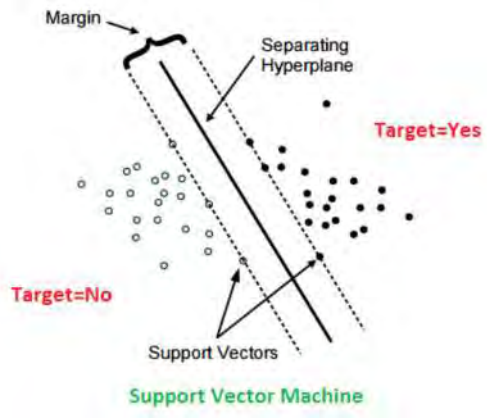


Figure No. 7: Support Vector Machine

VI. Results and Discussion

Actual Results:

Implementation Details

The implementation plan for an IPL match win-loss prediction using neural networks is as follows:

1. A database of all the IPL matches containing various match statistics is stored in MS Excel.

2. The user will be given a form that will be generated using a suitable programming language. This form will contain various attributes that will be used as testing data.

3. All the coding will be done in Python.

4. The Excel files are imported to Python, where the initial Excel database is taken as Input and later as target.

5. Training is performed on 70% of the data and testing on the remaining 30% of the data.

6. Initially, random weights are assigned all the entries in the database.

7. The weighted sum of all input nodes is then calculated and the activation for the node is calculated.

8. Necessary adjustments will be done according to error difference obtained between target output and achieved output.

Figure No. 8: Team Selection

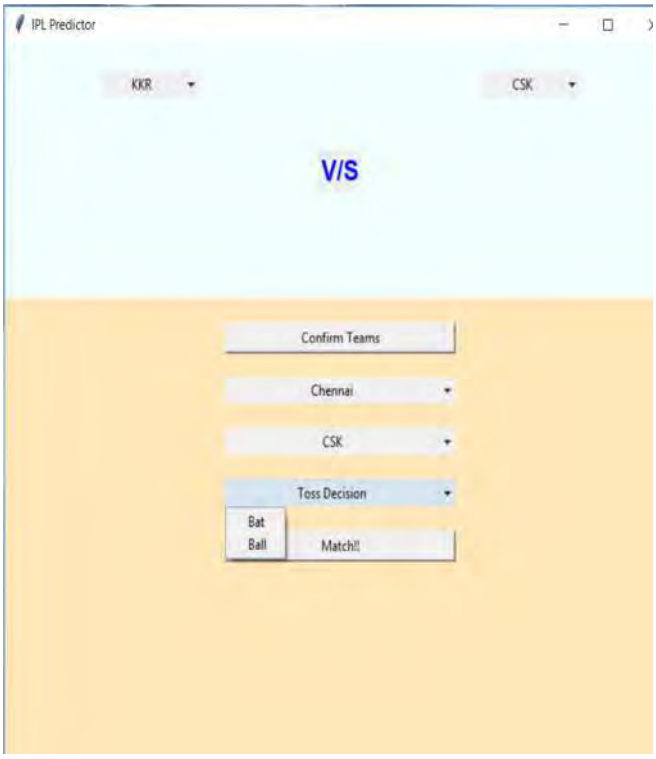


Figure No. 9: Toss Decision

Figure No. 10: Match Winner and Winning Team Score Prediction

Outputs:
We have used accuracy as a metric to evaluate correctness for all the algorithms.
Win-Loss prediction:

We obtained following accuracy for –

1) Support Vector Machines: 69.08%
2) Naïve Bayes Classifier: 64.70%
3) Random Forest Classifier: 51.55%

The algorithm with the maximum accuracy was used to predict the winning team for an IPL match.

Score Prediction:

Accuracy of neural network: 62.02%
Using the above model with given accuracy, the score for the winning team was predicted.

Outcomes:

The project intends to correctly analyse the input data on the basis of 14-16 parameters to correctly judge the winning team and its score. It uses complex algorithms like Naïve Bayes, Decision Tree, and Random Forest.

Discussion:

AI systems are now a part of a variety of specializations from stock trading to the setting of real estate prices;

from fraud detection to translation of vernacular languages; from designing our daily shopping lists to forecasting which movies we might like and diagnosis of medical condition.
This is merely the beginning. Not long from today, AI will be used to advance our understanding of biological phenomenon through analysis of highly dense and huge datasets, help us invent new drugs and make treatments more thorough and personal. Self-driving vehicles can revolutionize transportation and allow new paradigms in planning of architecture. Machines will automate our homes more efficiently, make businesses more productive and help predict risks to society.
There are several techniques that can be used to correctly classify and predict when a new tuple is entered.

VII. Future Scope:

Using neural networks and machine learning our project aims to deliver an application to predict the success of a team in an IPL match. By using neural networks and machine learning our project can be extended to develop a world cup-based application or for any other international sports tournaments.
To predict the success of the team we are taking the statistics of every innings of batsman as well as the bowler through every IPL season from 2008-2018. Also, the team statistics at a particular venue are considered. The prediction will be based on the outcomes to all the IPL matches in the past 11 years.
Since neural networks work on experience, we will be providing more than 600 different matches to our network, each with complete ball by ball data. This will help to gain accurate results. The number of matches won with a particular team combination, the average score of the team as well as the progress of the team through the season will be the deciding parameters.
Future research in this project can be predicting the bowl by bowl outcome of a match while the match is being played. This will require real time data collection as well as statistical measures for successful prediction.

VIII. Deployment:

Deployment can be a major challenge in software which involves machine learning. Huge amounts of processing is required just to get it up and running. Data generated in the process is humongous as well.
The ML model will be deployed using Google firebase or spring. The cloud platform will provide a base for future developments where flexibility, portability and reliability are required.
The GUI by itself is not very difficult to deploy. We have multiple ways of offering interfaces, such as web interfaces and command line interfaces. It will expand to a desktop GUI in the future as well.
The web interface will be deployed using Heroku or AWS, depending on the relative cost of hosting.
The command line can be used as a windows executable file. Users can simply download it and run it natively on their windows PCs.

IX. Conclusion:

By observing recent trends and strategies, the scope of machine learning is growing day by day. The project based on machine learning will evaluate and improve the understanding by using this project as a path. By using the technologies like ml, python, neural networks. Using our derived set of features, we will be able to predict the outcome of a match with approximately 65% accuracy, and the score with 60% accuracy. Predicting the outcome as well as the score of a cricket match has multiple factors involved, some of which we have tried to incorporate, others being difficult to correctly specify and represent in this project. Greater prediction accuracies can be achieved through more define data that better correlates to features significant for an IPL Match.

References:

[1] Duckworth, Frank C., and Anthony J. Lewis. "A fair method for resetting the target in interrupted one-day cricket matches." Journal of the Operational Research Society 49.3 (1998): 220-227.

[2] Beaudoin, David, and Tim B. Swartz. "The best batsmen and bowlers in one-day cricket." South African Statistical Journal 37.2 (2003): 203.

[3] Lewis, A. J. "Towards fairer measures of player performance in one-day cricket." Journal of the Operational Research Society 56.7 (2005): 804-815.

[4] Swartz, Tim B., Paramjit S. Gill, and David Beaudoin. "Optimal batting orders in one-day cricket." Computers and operations research 33.7 (2006): 1939-1950.

[5] Norman, John M., and Stephen R. Clarke. "Optimal batting orders in cricket." Journal of the Operational Research Society 61.6 (2010): 980-986.

[6] Kimber, Alan." A graphical display for comparing bowlers in cricket." Teaching Statistics 15.3 (1993): 84-86.

[7] Barr, G. D. I., and B. S. Kantor." A criterion for comparing and selecting batsmen in limited overs cricket." Journal of the Operational Research Society 55.12 (2004): 1266-1274.

[8] Van Staden, Paul Jacobus." Comparison of cricketers bowling and batting performances using graphical displays." (2009).

[9] Sankaranarayanan, Vignesh Veppur, Junaed Sattar, and Laks VS Lakshmanan." Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction." SDM. 2014.

[10] Khan, Mehvish, and Riddhi Shah." Role of External Factors on Outcome of a One Day International Cricket (ODI) Match and Predictive Analysis."

A GAN based Autoregressive approach to Music Generation

Aniket Ladukar, Advait Maduskar, Shubhankar Gore, Neha Patwari
Thakur College of Engineering and Technology, Mumbai, India
aniket.ladukar@gmail.com, advaitmaduskar.1@gmail.com, shubhankar27gore@gmail.com, neha.patwari@thakureducation.org

Abstract – **Music Generation is a challenging task from the perspective of machines. Even as humans, composing musical sequences is demanding to those that lack the knowledge of the nuances of music. Autoregressive models are effective for generation of locally coherent audio, but break down when it comes to generating globally coherent structures. In such cases, Generative Adversarial Networks, or GANs, prove to be substantially efficient if provided with sufficient data. The only drawback of GANs is the computational power required for the generation of musical structures. In this paper, we explore the applications of autoencoders as well as GAN networks for the generation of medium-to-long musical sequences** [1]**.**

Keywords – Generative Adversarial Networks, Autoregressive, music generation, deep learning, music

I. INTRODUCTION

Audio synthesis is very instrumental process in vast gamut of applications like music generation systems, audio synthesis algorithms used in text to speech conversions also known as synthesizers in music.

Synthesizers have a long history of being handmade devices taking a range of inputs. The inputs can be categorized as either control parameters (‘pitch’, ‘velocity’, etc..) or filter parameters (‘tone’, ‘timbre’, etc.). These inputs are then used to generate fine-tuned audio signals which we perceive as music.

Despite their drawbacks, or maybe due to them, synthesizers have had a deep impact on the advancement of music and culture in the past half century (Punk, 2014).

Music, like other diverse forms of information, has an elemental pattern indistinct to human eyes. Humans create new pieces music based on years of experience and instinct, rather than following some predefined mathematical formulae.

The objective of this project is to imitate a human like “thinking” in a computer. To achieve such a goal, the approach of using deep learning can be effective.

We aim to train an autoregressive model using a GAN based approach on current musical data to discern fundamental pattern in order to generate new audio sequences of medium to long duration.

II. BACKGROUND

One of the major issues with music generation is the lack of consistency at global and local level. Primitive papers written by Chen et al discuss about model that can create a melody but lacks harmony.

This proposes the following avenues to advance upon -

1. Music should be generated with proper tempo, variety of notes and various musical structures.
2. Create a model capable of producing music with long- and short-term coherent structures.

Liu et al tackle the same problem but are unable to overcome either challenge. They mention that their representation of music cannot accurately interpolate between the melody and the other parts of the piece, while also failing to note the complexity of musical pieces.

Two LSTM networks are used by Eck - one to process chord structure and local note structure and the other to learn long term dependencies in order to try to learn a tune and while maintaining coherence. However, this architecture is trained on set chords and thus cannot create a more diverse sequence of notes [2].

III. SIGNIFICANCE

Music can be considered as a complicated yet exceptionally structured continuous set of data. These structures can be found at timescales spanning from milliseconds all the way to the entire composition lasting several minutes. Creating such sequences consisting of temporal correlations arising from these structures is a demanding task because they span many different orders of magnitude.

The scope of this project is confined to conquer the challenge of music generation. It serves as a tool for taking individual skill a step ahead and speeds up composition of music. Moreover, the project will work as a tool to identify cases of plagiarism and settle cases of copyright infringement in the music industry. Ultimately the project highlights the cultural shift that is a result of incorporating deep learning with one of the bedrocks of human culture – music.

IV. OBJECTIVES

Like any other form of information, musical sequences have underlying patterns that are indiscernible to human eyes. Humans compose new music based on experience and intuition and not mathematical patterns. Our objective is to replicate this kind of "thinking" in a computer and to that purpose, the deep learning approach can prove fruitful. We aim to train a generative network on existing musical data and learn patterns in the data in order to generate new sequences of medium to long length.

We are working towards hypothesizing a new model for music generation that combines the properties of models we previously studied while omitting the shortcomings and integrating the positive aspects. We aim to conceptualize a model that amalgamates the properties of Autoencoders and GANs in a recurrent feedback loop.

V. METHODOLOGY

Classical music is being considered for the training of models for the following reasons:

1. MIDI files for classical music are available in abundance.
2. Classical music is more graded and regulated when compared to other genres.
3. It contains complex yet well-defined structures as opposed to the abstract nature that one might encounter in other styles of composition.

The files will be processed using the music21 library available in python. Each piece of music is decomposed into a sequence of chords with each chord having a predefined time period.

A chord is created for every corresponding new note. Each chord is represented as an encoded vector for processing. Almost 390,000 notes will be generated from a total of 700 A 70-15-15 split will be used for training, validation and testing respectively [3].

Both Generator and Discriminator are implemented as CNNs. The Generator develops along the time axis followed by the pitch axis. The Discriminator wraps in the opposite direction. The Generator is retrained for every five training iterations of the Discriminator [4].

VI. PROPOSED SYSTEM

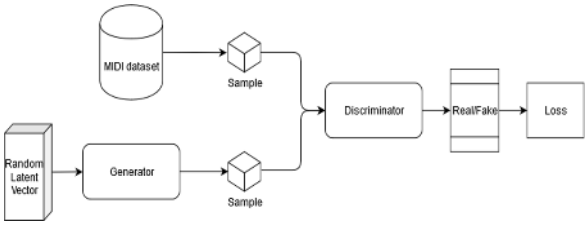


Fig 6. 1 GAN Architecture

The concept of GAN is fundamentally unambiguous and simple. The discriminator and the generator operate in an adversarial manner. The aim of Generator is to generate data which appears to be like the real data. The discriminator serves to distinguished the synthesized data from the genuine data. Here we plan to train the Generator once for every five training iteration of the discriminator. It is essential to realize that GAN is basically an approach taken to make two networks work together. Both Generator and Discriminator have their individual architecture [3] [4] [5].

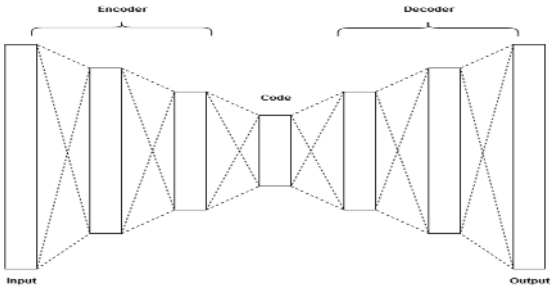


Fig 6. 2 Autoencoder architecture

Autoencoders are a distinct type of feed forward neural networks. Their basic goal is to reduce the dimensionality of the input. Here, the input and output remain the same. Autoencoders consist of mainly an encoder, a dimensional code and a decoder [6].

The encoder compresses the input into a dimensional code. This code is like a concise representation of the input and is also referred as the latent-space representation. The job of the decoder is to recreate the input from the dimensional code.

Autoencoders are very data-specific, i.e. they can only compress data similar to what they are trained on. The output generated by the autoencoder is not identical to the input. There is a certain amount of loss due to compression [6].

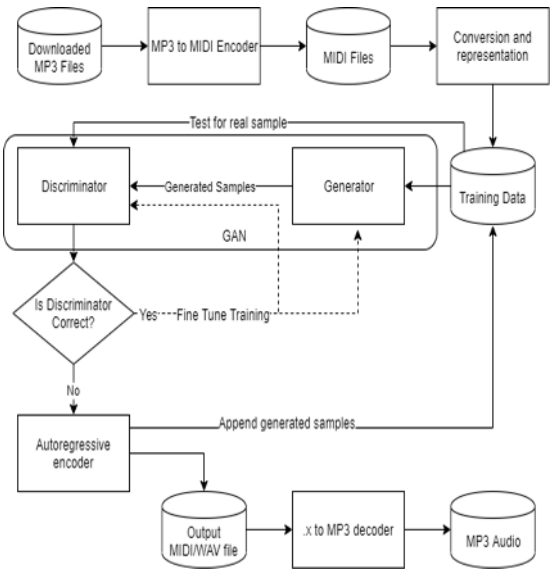


Fig 6. 3 Proposed System Architecture

Finally, we propose a system which combines both GAN and Autoencoder to generate using. By doing so, we can we aim to improve on the shortcomings of the existing models. Current models face one of the following issues –

- Training models which can generate quality results is a computationally heavy task.
- The models that train in a short span of time generally suffer in the quality of output that is produced.

The GAN based Autoregressive approach can prove to be fruitful because GANs being very effective in modelling an underlying pattern will produce quality results. Meanwhile, Autoencoders can be used to compress the data without losing important features will aid in speeding up the training process. This will eventually result in a faster and better generation of music.

VII. RESULTS ACHIEVED

A GAN based model was used to achieve the following results. The model was trained on a MIDI file of Al Adagi's composition. The pitch of the notes was the characteristic that was used to train the model. The generator learns to generate new audio sequences of varying timbre to represent different instruments.

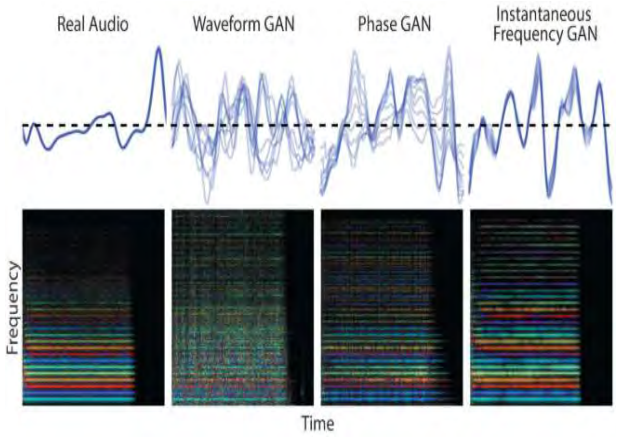


Fig 6. 4 Representation of musical waveform of musical waveforms

Here, we have used random interpolation to generate music on a scaled GAN architecture. After combining the MIDI files and parsing them, they can be represented as follows –

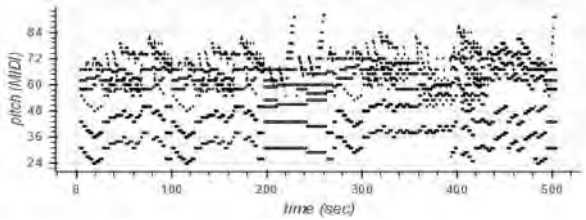


Fig 6. 5 Pitch vs Time representation of dataset used for random interpolation

Following the training of the model using this data, the time interval for random interpolation between instruments was set as 5 (sec/instrument) and constant Q-spectrogram of the generated music was obtained.

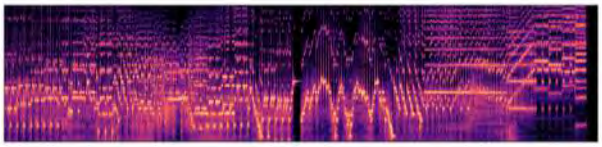


Fig 6. 6 Constant-Q spectrogram of generated music

Thus, from our dataset, we were able to randomly subsample, interpolate and generate 656 samples.

We observed that the interpolation between latent vectors was coarse and could be improved by increasing the time period between interpolations. Thus, to produce music that sounds smoother, we can set this time period to higher values and compare with currently produced music. Furthermore, more training data will help refine the model and help towards smoother conformation between instruments.

Few important observations include:

1. The tracks play at a analogous scales.
2. Some samples exhibit chord like playtime.
3. The bass plays at a lower pitch and is consistent in nature.
4. The drums tend to follow a balanced and musical pattern of 8 or 16 beats.
5. The pitch of instruments like piano and guitars tends to overlay each other with suggests the existence of symphonic affiliation.

VIII. FUTURE SCOPE

In the near future, we are focusing on improving established baselines by manipulating training data and interpolation times. We also plan on conducting a comparative analysis with other methods for music generation with various permutation of functional parameters.

We aim to improve the musical and aesthetic appeal of the music generated by the model as it lacks certain qualities and falls behind the music generated by consummate artists.

Future scope includes development of an autoregressive GAN model in a recurrent feedback loop that can be used not just for music generation, but also for detection of plagiarism in music by introducing original work as a sample data point and the dubious work as the fake samples in the discriminator function of the GAN.

References

1. J. A. Henning, A. Umakantha and R. C. Williamson, "A classifying variational autoencoder with the application to polyphonic music generation," arXiv preprint arXiv:1711.07050, 2017.
2. I. Malik and C. H. Ek, "Neural translation of musical style," arXiv preprint arXiv:1708.03535, 2017.
3. H. W. Dong, W. Y. Hsiao, L. C. Yang and Y. H. Yang, "MuseGAN: demonstration of a convolutional GAN based model for generating multi-track piano-rolls.," in International Society of Music Information Retrieval Conference, 2017.
4. H. W. Dong, W. Y. Hsiao, L. C. Yang and Y. H. Yang, "MuseGAN: Multitrack sequential generative adversarial networks for symbolic music generation and accompaniment," in Third Second AAAI Conference on Artificial Intelligence, 2018.
5. I. P. Yamshchikov and A. Tikhonov, "Music generation with variational recurrent autoencoder supported by history," arXiv preprint arXiv:1705.05458, 2017.
6. J. Engel, C. Resnick, A. Roberts, S. Dielman, M. Norouzi, D. Eck and K. Simonyan, "Neural audio synthesis of musical notes with wavenet autoencoders.," in 34th International Conference on Machine Learning, 2017.

Integration of Hyperledger Fabric in SDLC

Raj Mandal, Gopi Mehta, Aditya Nambiar, Sandip Bankar
Department of Information Technology, Thakur College of Engineering and Technology, Kandivali (East), Mumbai, India – 400101
rajmandal800@gmail.com, gopimehta60@gmail.com, aditya.nambiar24@gmail.com, sandipbankar1@gmail.com

Abstract – **Software development methodologies are needed so that the software development life cycle can become systematic to complete application development within the right time frame. This paper focuses on architecture that redefines the DevOps pipeline by turning the Version Control System (VCS) and the Integration Server (IS) into a decentralized system as a separate and independent fabric channel. The advantage of having such a system is that it prevents the problems occurring in the pipeline due to centralized architecture. This also allows the system to generate reports at each stage hence improving the compliance and transparency of the architecture. The report consists of well-formed data that is derived from transactions up until that point thus providing the ability to monitor concurrently the progress at any stage and at any time.**

Keywords: Hyperledger Fabric, SDLC, Blockchain, version control, decentralization, DevOps, Continuous Monitoring, Continuous Integration.

I. INTRODUCTION

One of the most fundamental notions of the software development process is SDLC models which stand for Software Development Life Cycle models. SDLC–is a continuous process that begins at the moment when a decision is made to launch the project and ends at the moment when it is permanently removed from exploitation [1]. No single model exists with an SDLC. It is divided into major groups, each having its own features and weaknesses. They originated from the first and oldest SDLC "waterfall" method, and their range grew significantly. The scope of SDLC models is dictated by the wide range of product types-starting with the development of web applications to complex medical devices. The problem with 'Agile methodology' came into picture when things started to get difficult for the organizations who took up client services. Development teams tended to deploy their work on a large scale onto the operations team. There was very less coordination between the teams from all levels, specifically teams that were responsible for developing the service (development team) and the teams that were responsible for bringing the service into delivery phase / production phase onto the customer's hand (operations team). The operations team was not on the same pace as the development team. So for some organization work standards, agile methodology was not enough. DevOps is an IT Culture [2]. It consist of various practises and the approaches that influence organizations in organizing teams, build systems and influence even the structure of the systems that you create. Hyperledger Fabric is a Blockchain development framework and one of Hyperledger's collaborators in the Linux Foundation. Intended as a platform for software development or modular architecture implementation, Hyperledger Fabric enables plug-and play applications, such as consensus and membership services. Hyperledger Fabric leverages container technology to host "Chaincode" smart contracts that exemplify the application logic of the system. Originally,

II. MOTIVATION

Multiple entities are involved in the overall software development life cycle (SDLC) throughout the traditional approach to application development and maintenance [1]. This approach is often granulated in terms of features, cost, schedule, quality, performance, and other administrative overheads such as vendor integration by techniques, frameworks, processes, people, and tools that impact the final product. The SDLC operations and production engineers work together throughout the project lifecycle from design through to the development process to assist in the whole production.

It is therefore important to bridge the gap between engineering and operations team by establishing a meeting point for two of Blockchain and SDLC's most promising technologies. Blockchain's strength lies in its decentralized nature [4], which means that each member of its peer-to-peer network can depend on the other. If some Blockchain businesses have systems that are optimally incorporated with the

system and can produce high-quality software, but others cannot, the trust in the system begins to break down, and those within it. Implementing Blockchain with a DevOps approach is an IT philosophy that can prevent this issue. This consists of different practices and strategies tha

t affect organisations in coordinating teams, building systems and affecting even the program structure y ou are building [2].

III. EXISTING SYSTEM

Recent work is moving from traditional software development models to DevOps approaches. DevOps came into existence because of certain problems in the traditional way of managing development and operations. DevOps is mode of work, by using a series of tools to increase the communication and collaboration between development teams and operations teams to reduce the conflicts between the two teams and improve

the development efficiency and quality. This section will give an overview of the main challenges [6, 7] that sparked the evolution of proposed method.

- Lack of trust between Development and Operations
- Communication, Collaboration gap between development and operations
- When problem occurs, people may blame between teams
- Geographical distribution can create challenges
- Lack of management structure between the developers and the operations
- DevOps has not been systematically organized and managed in large-scale projects
- The working pressure on the operations team is growing
- Lack of senior management involvement, stubbornness mindset.
- People is a challenge when using DevOps, as people have to work together in the ways they never saw before
- Developer and Operations lack of interest in the "other side"
- Change in mindset to achieve successful DevOps is a challenge
- DevOps artifacts are usually bound to certain tools, which make it difficult to reuse
- Different software versions and dependencies
- Smoothly deliver code from development to production is a challenge

The reality of Software Development Life Cycle is very complex and untrusted in enterprise with multiple lines of business. In an environment of rapid software development and deployment it is necessary to strengthen the DevOps by providing decentralized data storage of documents on Blockchain to increase trust, auditability and traceability.

3.1 Continuous Monitoring and its drawbacks

Today organizations are required to deliver greater levels of customer satisfaction through their online services. Many, however, are forced to support these initiatives with an interrupt-driven approach as they react to repair things after they break. However, for a more proactive approach and to manage expected high levels of SLAs, organizations can prefer to reduce their amount of unscheduled downtime by implementing Continuous Delivery (CD) model to their development strategy and add to their efforts. What's critical within the CD model is that the ability to watch and manage systems using a structured way, getting early detection of problems so organizations can make changes before the system is impacted. While unexpected failures make interrupt-driven work unavoidable sometimes, organizations can become more proactive by examining their current approach and toolset against business has to help create a path to continuous service delivery optimization.

DevOps involves a lot of tools and processes that go within an organization's pipeline. Monitoring of these tools play a major role in management of these tools [2]. When an organization automates their pipeline, namely implementing a Continuous Integration / Continuous Deployment (CI/CD) model for their software engineering process, they need to be also ready for the different problems that can occur at each stage of the pipeline. Generally, companies tend to use tools like Nagios or rely upon Cloud services for their deployment procedures but using a tool also requires you to itself manage that tool itself since they are licensed and versioned preciously. Companies cannot afford to spend time looking after the tools when they got their own product deployment to worry about.

Most software organizations lack the idea of monitoring their pipeline from start to finish in an environment of multiple somewhere in the pipeline. Monitoring solutions that companies are trying to adopt can become quite complex when they have numerous microservices to look after [2]. A Continuous Monitoring solution is presented as a better solution to solve this issue of having to track numerous microservices from previous stages. On a second note, Kubernetes which is a popular Orchestration service becomes complex enough that it requires a separate set of operations microservices, where their product is ready or stuck engineers to manage the tool and again, this tool comes with its own set of licensing because the company chose to use an external tool and does not have anything they can use within their infrastructure.

IV. Proposed Solution

In the traditional SDLC models usually, there are separate environments and teams at every stage of the delivery process, each separated from the other by physical and logical controls, which slows down the delivery of the final product. So, there is a need for a solution to speed up the overall DevOps lifecycle as well as control the separation of duties in an agile environment. Blockchain has been one such technology that has proven its advantages such as recording legitimate transactions, traceability of origin of an entity, enhanced security model, transparency. Software industries expect newer technologies to overcome the hurdles they face after implementing legacy technologies in an out-of-the- box manner. Therefore, Blockchain technology can be leveraged to build more open and secure software distribution chains. By using Blockchain technology, specifically its traceability feature, we can not only develop better software delivery chains but also eliminate the usage of

third-party tools such as continuous integration servers and automated test suite to trace the successful movement of code through the pipeline. With the help of Blockchain's various advantages, monitoring different stages in a DevOps pipeline becomes natural. Companies do not have to look out for better tools now that a technology provides a solution for Continuous Monitoring. The architecture focuses on relying on the Blockchain data structure to provide more of a timeline oriented architecture for monitoring each stage in any DevOps pipeline irrespective of what technology is used to perform the tasks during any stage.

4.1 Features of Hyperledger Fabric framework

1. Permissioned membership - The Hyperledger Fabric is a Blockchain system for permitted networks. For each network member this structure identifies a certain identity. Access control policies are enforced which require authorized access to network participants.
2. Performance, scalability and levels of trust - Hyperledger Fabric is a modular architecture that divides the transaction processing into threephases: distributed logic processing and agreement (chain code), ordering of transactions, and validating and committing transactions. This separation or modularity improves the scalability of the network and requires fewer verification procedures across the nodes leading to optimized performance.
3. Modular architecture supporting plug-in components - Hyperledger Fabric's modular architecture lends developers leverage to use their preferred components. Many multi- company networks, for example, have their own identity management systems which can be used as a plug-in to rebuild.
4. Rich Queries - By default, Hyperledger fabric uses LevelDB database for the state management. This database provides only simple query support but Hyperledger Fabric can be integrated with CouchDB, a database that provides rich query support. Thus, complex business logic and queries can be implemented for the advanced systems.

4.2 Utilizing critical features of the framework

The proposed architecture is currently focused on utilizing few specific advantages namely, Permissioned membership and Channels. Permissioned Membership allows us to audit each transaction and members associated with the transaction and Channels help in isolating and implementing a separate audit for a few important parts of the pipeline by having a separate copy of ledger. This framework helps our product to perform parallel monitoring throughout the lifecycle of a product. The transactions [1] that are present in our Chaincode / Smart Contract in This work are based on all the inputs and outputs of various tools and participant's documents.

V. Proposed Architecture

5.1 Architecture Overview

This work consist of a decentralized architecture. The proposed solution redefines the DevOps pipeline by making Version Control System (VCS), Integration Server (IS) as separate, isolated Hyperledger Fabric channels feature. Isolation also allows us to make our pipeline have a pluggable environment.

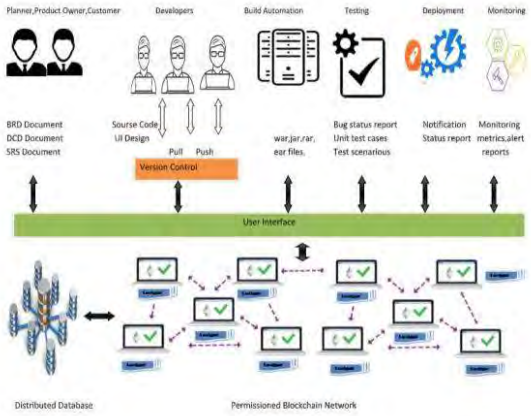


Fig. 5.1: Project Architecture

Figure 5.1 shows the overall architecture which covers the different participants, inclusion of the Distributed Database (IPFS) and Permissioned Blockchain Network (Hyperledger Fabric).

5.2 Role of DevOps

This proposed system consists of a deployment pipeline that permits parallel monitoring service and storage in a distributed database. This work focuses on preventing the problems of having a centralized architecture in the pipeline. Furthermore, keeping the two channels locally available will not disrupt the development team from committing their code and checking whether it fits with all the other modules coming from other teams built into one executable application.

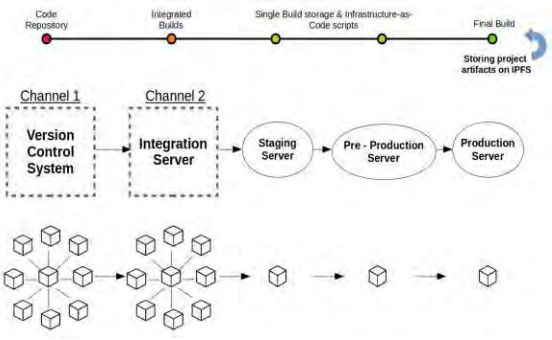


Fig. 5.2: Proposed DevOps pipeline

Figure 5.2 shows the proposed DevOps pipeline which is flexible as it does not depend upon any certain tool that

an organization would like to use or using in their current development or operations stages.

The architecture currently plan on allowing Operations side to be hosted on the internet since it involves provisioning real-time Virtual Machines (VMs) and that has to be done physically. However, the scripts they write for provisioning as seen in case of tools like Chef or Puppet configuration management, will be going under our VCS channel.

Compliance and transparency throughout each stage of DevOps pipeline.

The ability to perform Continuous Monitoring and able to generate reports at each stage of the pipeline is not only important to the people working on building and maintaining the product but also to the people who proposed the work on this product. They have the right to see the status of their product and decide the compliance value of an organization. This will also add to the key point that organization is being as transparent as it can to satisfy their customer's needs and making sure they are regularly keeping their clients updated about the status of the product they proposed.

5.3 Role of Blockchain

This work enables transparency throughout the pipeline. Compliance specialists are able to fetch a summarized report of all the relevant data that can be flexibly added.

Continuous Monitoring is also enabled throughout the pipeline as the underlying Blockchain is storing all the transactions such as data entering / passed to a tool at each stage in pipeline and similarly recording the output from each tool. Thereby ensuring that each transaction that occurred can be traced back to know relevant details about a certain stage of the pipeline.

5.4 Role of Interplanetary File System (IPFS)

The backend file store, namely, IPFS file store allows us to store files on IPFS nodes set in a network or in a cluster of nodes on the network and provides a way to access these files rapidly without huge latency that was caused due to dependency on a far-away server that hosts the file. So even when a node or peer who is situated physically next to us has the file you desire, the network fetches the file contents from that node and does not rely on the central server. Nodes communicate with each other and they are able to store the files of each other. They only store the files that are requested by the nodes themselves.

IPFS allows accessing files by their metadata's hash. It is termed as a "multihash". The purpose is to remove the dependency from referring to a physical location (via IP addresses) and referring by content- based addressing mechanism to allow looking up on any peer that is in the network and fetch the desired file or folder faster as compared to a server's response timeRelying on the hash of file / folder (collection of files) is that one is able to preserve the integrity of the file uploaded. Any change to the file will result in a new hash of the file and would require re-uploading the file and thus a new reference is generated to be able to access this modified file.

All the artifacts that are uploaded to or produced after a version control, build or testing tool's operations will be stored on IPFS. IPFS utilizes content based addressing and allows us to fetch files with the help of file hash, from a nearby IPFS node. This removes the unwanted overhead of relying on the URL based addressing mechanisms if an organization were to go for cloud based DevOps pipeline solutions.

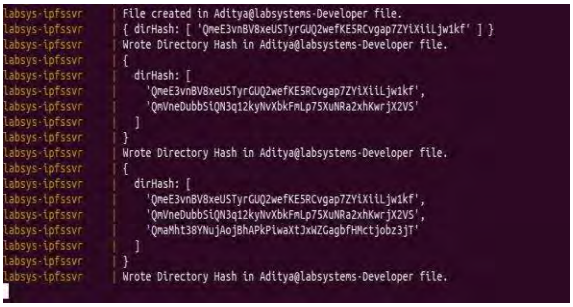


Fig. 5.4: IPFS content-based addressing.

Figure 5.4 shows the way this work stores hashes of different directories in a JSON file. It allows us to keep track of the different versions irrespective of whether it's a single file or a batch of files since any uploaded file is stored in a directory named 'DocumentsDir'. The 'dirHash' JSON array stores different versions of the directory hash as and when the files within the directory are modified or new files are added.

The files or artifacts are stored in the form of chunks of 256kb. After which it utilizes a Linked List based referencing mechanism to link chunks of a relatively larger file. This means that users need not worry about confidentiality of their project artifacts.

VI. IMPORTANCE OF THE ARCHITECTURE

Version control is essential to documents that undergo a lot of revision and redrafting, and as a result of this is particularly important for electronic documents, they can simply be updated by a number of users. These changes might not be directly apparent. Knowing what version of a document you're starting at is vital, for instance, if you're making an attempt to search out that version of a policy is presently good, or that version of the policy was in use at a selected time. Version management is additionally vital if you're engaged in a cooperative document with a variety of contributors and/or with frequent revisions. As you build and implement documents such as standardized rules or project documents version

tracking, you can find the document occurrence over a prolonged duration, e.g. multiple versions of a draft and final regulation. This allows traceability-retention of drafts and details of changes created, typically as a result of contributions from multiple collaborators within the development of documents, order of changes, and recording of those versions of documents considered final and/or approved by relevant teams or individuals.

It has the potential to link documents to choices, contributions, contributors, and time. All documents are associated by metadata such as computer filename, author, and date. Multiple versions of documents are also to be differentiated, so the latest version is known. Also, continuous integration is the practice of combining feature branches into the main branch of code to be designed and tested mechanically. Since this is often an automatic flow, this could be done anyplace from many times daily to many times per week.

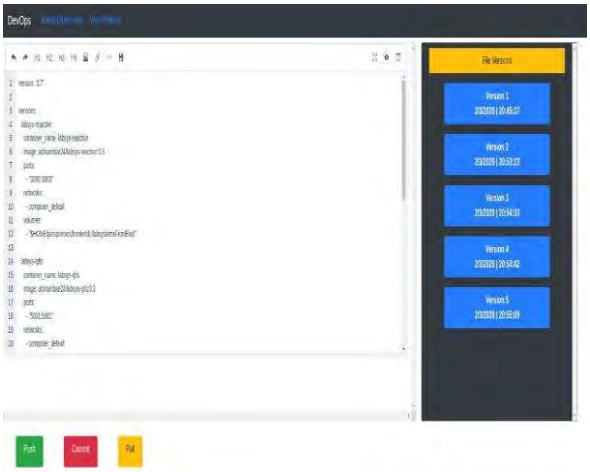


Fig. 6: File versions

Figure 6 shows the different versions of files that a developer handles. These versions are generated whenever a 'Commit' operation is executed by clicking on the Red 'Commit' button on the User Interface. Also user has to push their changes before commit since the idea is to only create versions of the file where you will be having significant changes in each version you view for the file.

The utilization of a git repository acts as a trigger for the build server to compile, execute and, take a look at the code changes. If code changes cause the automatic take a look at cases to fail, associate alarm are raised to tell the developers that their amendment was a failure. The developers can then either rollback their changes to the antecedent productive version of the code or roll forward to bring in a fix to the code thatcaused the failure. Continuous Integration / Continuous Deployment (CI / CD) is the agile development lifecycle which is beneficial to large organizations compared to the normal waterfall where the delivery part would last weeks to months after integration and testing phases.

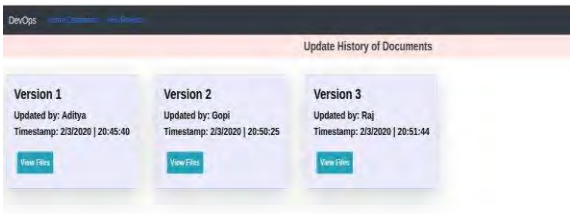


Fig. 6.2: Users' project timeline

Figure 6.2 shows the different versions of the batch of files, we named it as 'Documents'. The Figure depicts the usage of the auditability and traceability of the batch of files a participant uploads. The architecture allows another participant (peer) to update a document of other peer since the fundamental feature of Blockchain is a distributed ledger which allows all participants to have access to it and perform transactions and record on it.

VII. Future Scope

Machine learning is a technology that now everyone is aware of as it has changed the entire IT industry because of the capabilities it possesses and the ability to give more productive output. Going on the similar lines we look forward to integrate the power of Blockchain with machine learning alongside DevOps. The proposed solution will use reinforcement learning to predict the possible interruptions in the development pipeline and most importantly we will train the model in such a way that it automatically detects any problem arising in the pipeline. Furthermore, the model will learn over the period of time through its experience and prepare its own knowledge base so that in future it can figure out any problem in the pipeline more efficiently.

VIII. Conclusion

The purpose of using Hyperledger Fabric in the architecture was clearly to do justice to the corporate standards. This includes the organizational disciplinary behaviors which is incorporated by the Hyperledger Fabric framework quite well. Its features enable one to utilize it as a base for development in the corporate environment. This also makes sure that the organization tends to cope up with the trending and upcoming technologies while not leaving behind its traditional principles. The practice of SDLC in an organization tends to help in deploying the outcome of a project faster by including the Operations team with the development team which is quite different than what is done traditionally in the real world. This also ensures the product quality and also withstands most of the flaws and potential rectifications that must be made once the product passes various tests and eventually the customer acceptance test. The purpose of this architecture was to merely build the SDLC practices

onto the Hyperledger Fabric platform such that organizations find it easier, better and still maintain the same quality of the product as they would without any additional functionalities.

IX. 9. ACKNOWLEDGEMENT

The quality and ultimate outcome of this paper involved a great deal of support and expertise from many people and we are incredibly privileged to have successfully completed the paper. We take this opportunity to thank everybody for their support.

We sincerely thank our principal Dr. B. K. Mishra, vice Principal, Dr. Deven Shah, Mentor Dean Dr. Kamal Shah and HOD Dr. Rajesh Bansode for always guiding us to paths that were right for us. We are grateful to our guide Mr. Sandip Bankar who supported and constantly monitored this work and helped us in not only completing this project but also providing us with basic knowledge with their practical aspect that really proved beneficial to us. We are forever thankful and fortunate enough to receive constant encouragement, support and guidance from all IT department faculties that helped us finalize our paper work successfully. We would also like to extend our sincere gratitude to all the laboratory support staff for their timely support. We would like to convey our heartfelt thanks to our parents for their kind cooperation and encouragement without which none of this was possible.

References

1. Mohit Kumar Sharma (2017) A study of SDLC to develop well engineered software, International Journal of Advanced Research in Computer Science
2. Blockchain in DevOps. (2020). [ebook] Ernst & Young, p.10. Available at: https://as sets.ey.com/content/dam/ey- sites/ey-com/en_ca/topics/blockchain/ey-blockchain-in- devops.pdf [Accessed 4 Feb. 2020].
3. Bass L, Weber I, Zhu L (2015) DevOps: A Software Architect's Perspective. Addison-Wesley, Old Tappan, New Jersey.
4. Andreas M. Antonopoulos (2014) Mining and Consensus. In: Mastering Bitcoin: Programming the Open Blockchain, 1st edn, O'Reilly Media, Inc, California, United States. P 181
5. Bass L,Weber I, Zhu L (2015) Moving a System through the Deployment Pipeline. In: DevOps: A Software Architect's Perspective, 1st edn. Addison-Wesley, Old Tappan, New Jersey, United States. p 81
6. Mali Senapathi, Jim Buchan, Hady Osman (2018) DevOps Capabilities, Practices, and Challenges: Insights from a Case Study, EASE'18, Christchurch, New Zealand
7. Leonardo Leite, Carla Rocha, Fabio Kon, Dejan Milojicic, Paulo Meirelles (2019) A Survey of DevOps Concepts and Challenges, vol. 1, ACM Computing Surveys

Social Media Sarcasm Detection

Ashish Tiwari, Kshitij Singh, Manish Rai, Neha Patwari
B.E. Information Technology
Thakur College of Engineering
and Technology
Mumbai, India
tashish9711@gmail.com, Singhkshitij340@gmail.com, Manishraimnm97@gmail.com, neha.patwari83@gmail.com

Abstract- **Sarcasm is an advanced type of assumption articulation where speaker express their feelings inverse of what they mean. Sarcasm location and Emotion identification from long range informal communication destinations has been an extraordinary field of study. With the development of e-administrations, for example, web-based business, e-the travel industry and e-business, the organizations are excited about misusing feeling and sarcasm examination for their promoting methodologies so as to assess the open demeanors towards their image. Consequently, productive feeling and sarcasm demonstrating framework can be a decent answer for the above issue. This work targets building up a framework that gatherings post dependent on feelings, assessment and find mocking posts, if present. The proposed framework is to build up a model that help to go to an induction about the feelings of the posts specifically outrage, shock, cheerful, dread, distress, trust, expectation and disturb with three sentic levels in each. This aides in better comprehension of the presents when looked at on the methodologies which detects the extremity of the posts and gives only their suppositions i.e., positive, negative or nonpartisan. The posts dealing with these feelings may be snide as well. The Sentiment and feeling recognizable proof module distinguish the notion or feeling of the post by assessing score of each word in the remark which is utilized by sarcasm recognition techniques to identify sarcasm. The feeling recognizable proof module utilizes the lexical databases Word Net, SentiWordNet to locate the correct opinion scores for the words as for every feeling. It likewise utilizes Sarcasm discovery calculations like Emoticon sarcasm recognition, Hybrid sarcasm location, Hashtag Processing, Interjection Word Start (IWT).**

Keywords- SentiWordNet, Emoticon sarcasm detection, Hashtag Processing

I. INTRODUCTION

There are various patterns opening in the period of slant examination, which dissect 'mentality and assessment individuals in web-based life, which including social destinations like Facebook, Twitter, web journals, and so on. [1] The primary point of opinion investigation is to distinguish the extremity (positive, negative or impartial) in a given text. Sarcasm is a unique kind of assessment which can flip the extremity of the given content. [3] Sarcasm is characterized as 'the utilization of incongruity to taunt or pass on hatred'. Sarcasm is a complex type of conclusion articulation where speaker express their assessments inverse of what they mean. Sarcasm is a complexity between positive opinion word and a negative circumstance. What makes undertaking of recognizing sarcasm hard is that even people think that it's difficult to comprehend them now and again without earlier information on the point. Model: "Goodness! He out on a duck, what a legendary batsman". Right now, use to communicate the positive feeling (significance) however by and large tweet reflect negative conclusion toward the batsman. In contrast to a basic invalidation, a wry book, normally passes on a negative sentiment utilizing just positive word

II. BACKGROUND

Scarcely any innovation patterns of the previous decade have been more conspicuous than the ascent of Big Data. Driven by the accessibility of modest attractive stockpiling on a formerly remarkable scale, this bounty of information speaks to a tremendously important asset if significant data can be precisely removed. Information mining has gotten extensive consideration especially in the field of assumption examination which looks to extricate and characterize emotional substance. The use of ground-breaking factual procedures to customary normal language preparing strategies has yielded fascinating outcomes with regards to numerous zones. Notwithstanding, the acknowledgment and order of silliness, and specifically sarcasm, stays a difficult issue and to be sure it is one that structures the ideal test stage for a propelled degree of comprehension by a PC. Sarcasm has for some time been viewed as a fascinating issue with regards to phonetics, however right ID of its quality additionally speaks to a chance to accomplish exactness gains in the field of notion investigation. Since 2001, assumption investigation, a field of Natural Language Processing that worries about deciding the extremity of emotional substance, has seen significant research premium. This intrigue has been fundamentally roused by its conceivably important business applications. As exhibited by, the effect of online surveys on item buy is generous. Directed slant examination permits an organization.

Association to amplify their responsiveness to such audits by encouraging an evaluation of their online notoriety dependent on the proof of social cooperation's. Additionally, exact opinion examination permits a company to measure the adequacy of a promoting effort or to test the capability of a market opportunity, both significant vital capacities. From a

logical point of view, it has empowered people to look at human reactions to social wonder on a worldwide scale. Given these potential advantages, Twitter and other advanced stages has normally pulled in both corporate and scholarly consideration for leading notion examination as one of the biggest and most assorted internet-based life stages.

III. PROBLEM DEFINITION

Sarcasm sentences can be utilized nearly in all points. They can take variable linguistic structures. Likewise, to get sarcasm, one needs to know the setting of the sentence. For example, the sentence "I love being rich" isn't sarcastic without anyone else. In any case, on the off chance that you realize that the speaker is poor, you will conclude this is a sarcastic sentence. In this manner, to identify sarcasm, you must have earlier information about the subject or sarcasm, which may not generally be accessible.[5]

Sarcasm is defined as a cutting, often ironic remark intended to express contempt or ridicule. Sarcasm detection is the task of correctly labelling the text as “sarcastic” and “non-sarcastic”. It is a challenging task owing to the lack of intonation and facial expression in text. Nonetheless humans can still spot a sarcastic sentiment in the text and reason about what makes it.So, recognizing sarcasm in text is an important task for Natural Language Processing to avoid misinterpretation of sarcastic statements as literal statements.

IV. PROPOSED METHODOLOGY

We shall exploit a deep neural network for sarcasm detection, we shall build a neural model, with two sub neural networks to capture tweet content and contextual information, respectively. The two-component structure closely corresponds to the two feature sources of the baseline discrete model. We will model the tweet content with a gated recurrent neural network (GRNN) OR Long-Short-Term-Memory.

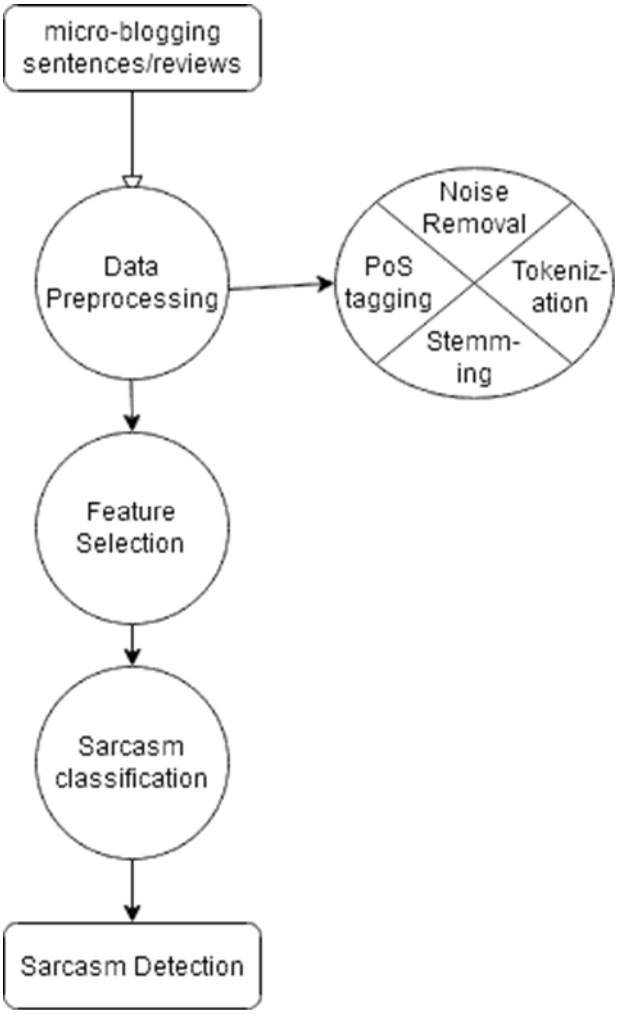


Fig.4.1 Methodology

Micro-blogging Sentences/Review: Micro –blogging sentences are extracting from various Social media like Reddit, twitter, Facebook etc.[11]

Data Preprocessing: In the data preprocessing stage at first, the info data is taken and hashtags are distinguished. These hashtags are expelled from the information data. This module incorporates Tokenization, Stemming, Noise Removal, and PoS Tagging.

Noise Removal: Noise Removal from the info information is the way toward evacuating the stop words, URLs, accentuation and so on. These are expelled from information with the goal that the purified information can be effectively prepared for the following stage.

Tokenization: Tokenization is the way toward parting the information into little units called tokens. The information is isolated into tokens in order to handily process the information.

POS Tagging: It is the way toward coordinating a word to its linguistic class, which assists with understanding its job inside the sentence. Basic grammatical features considered in Pos Tagging are Noun, Verb, Adverbs, and Conjunctions, and so on. Grammatical feature taggers for the most part accept a grouping of words as information and give a rundown of tuples as yield, where each word is related with the related tag.

Stemming: Stemming is the way toward shortening a word to its root structure. By utilizing the stemming method, the quantity of words can be diminished in the information. Porter Stemming algorithm is used for the stemming process.

Feature Selection: Feature Selection is the way toward separating the necessary features from the accessible info dataset. The necessary features resemble tweet id, tweet, date of the tweet and so forth.

Classification of a sarcasm: In the classification stage, the item surveys/remarks/tweets are gathered and are characterized utilizing classification calculations so as to distinguish the sarcasm in the tweets. For tweets, the classification is sure extremity or negative extremity. For the item surveys the classification procedure used to distinguish audit thinking about a size of 5.

Sarcasm Detection: It is the way toward distinguishing whether a tweet is sure, negative. Sentiment Analysis is useful for the advertisers to perceive the popular conclusion about their organization and items, and furthermore to think about consumer loyalty.

V. RELATED WORK

Semantic demonstrating of sentence significance is an all-around inquired about point in NLP.[3] Because of 'terrible language' in Twitter and a recognizable drop of precision for start of-the-workmanship electorate parsers on tweets, the semantic displaying of tweets has caught the consideration of analysts. To manufacture a semantic portrayal of a sentence in different NLP undertakings, for example, notion examination, analysts have utilized syntactic structure to form an all-out portrayal as a capacity of the word-vector portrayal of a sentence's parts. (Nakagawa et al., 2010) depicts a

Tree-CRF classifier which utilize an information driven reliance parser, maltparser2, to acquire a parse tree for a sentence, and whose arrangement work utilizes the head-modifier relations of the parse tree. (Mitchelland Lapata, 2010) and (Mitchell and Lapata, 2008) defined the sythesis capacity of a sentence by mathematical tasks over word meaning vectors to obtain sentence meaning vectors. (Guevara, 2010) and (Malakasiotis, 2011) defined their structure work utilizing a lot of explicit syntactic relations or explicit word classifications (Baroni and Zamparelli, 2010). (Socher et al., 2011) proposed an organized recursive neural system dependent on the convolution activity, while (Kalch brenner et al., 2014) proposed a convolution neural system (CNN) with dynamic k-max pooling, considering max pooling as an element of the info length. For sarcasm identification, because of the unpredictability of the undertaking and the to some degree more unfortunate exactness of beginning of-the-craftsmanship body electorate parsers on tweets, specialists have considered surface-level lexical and syntactic signs as genuine highlights. Kreuz and Caucci (Kreuz and Caucci,2007) investigated the job of lexical pointers, for example, additions (e.g., "hmm" or "gosh"), punctuation symbols (e.g., '?'), intensifiers, and different linguistic markers for example non-veridicality and metaphor, in recognizing sarcasm in stories. (Tsur et al.,2010) noticed the event of "whoopee!" or "incredible!" As a common part of snide examples in Amazon product surveys. (Davidov et al., 2010) analyzed the viability of online networking markers, for example, hashtags to distinguish sarcasm. Lukin (Lukinand Walker, 2013) proposed a potential bootstrapping technique for sarcasm grouping in social discourse extend lexical N-gram prompts identified with sarcasm (for example "gracious extremely", "no chance", and so on.) just as lexico-syntactic examples. (Riloff et al.,2013) and (Liebrecht et al., 2013) applied N-grams highlights to a classifier for English and Dutch tweets and saw that a few themes repeat regularly in mocking tweets, for example, schools, dental specialists, church life, open vehicle, the climate, etc

VI. NEURAL NETWORK

Semantic displaying of sentence importance utilizing neural systems has been an objective of consideration in the internet-based life network. Neural system structures, for example, CNN, DNN, RNN, and Recursive Neural Systems (RecNN) have indicated phenomenal capacities for demonstrating complex word piece in a sentence. A snide book can be considered essentially as a grouping of content signals or word blends. RNN is an ideal fit for displaying fleeting content signals as it incorporates a worldly memory part, which permits the model to store the transient logical data straightforwardly in the model. It can total the whole arrangement into a worldly setting that is liberated from express size limitations. Among the numerous usages of RNNs[3]

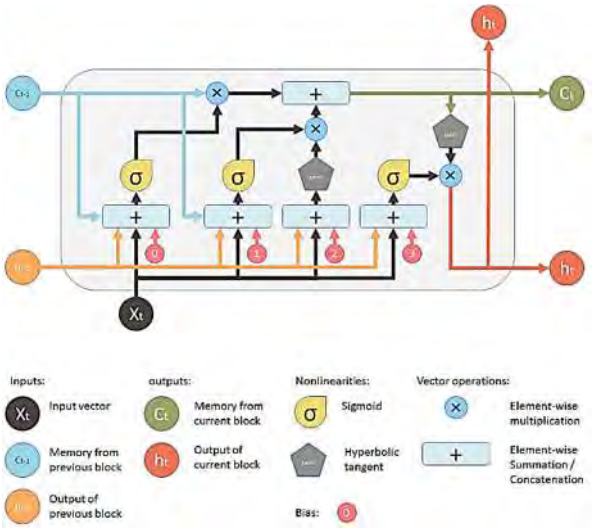


Fig 6.1Structural LSTM[9]

LSTMs are anything but difficult to prepare and don't experience the ill effects of evaporating or detonating slopes while per shaping back proliferation through time. LSTM has the ability to recall long separation fleeting conditions. In addition, as they perform transient content displaying over info highlights, more elevated level demonstrating can recognize elements of phonetic variety inside the information.

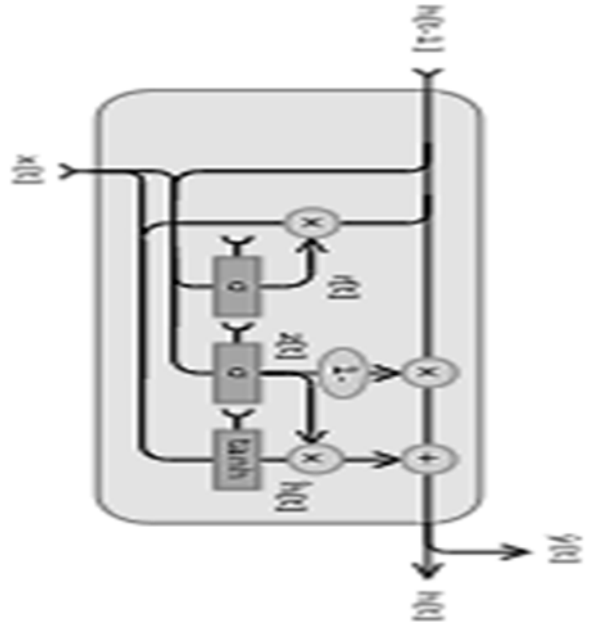


Fig6.2 Structural Gated recurrent unit[10]

RNN has shown the intensity of semantic displaying productively by joining criticism cycles in the system engineering. RNN systems incorporate a fleeting memory part, which permits the model to store the fleeting relevant data straightforwardly in the model. At each time step, it thinks about the present info xt and concealed state ht−1.[1] Thus the RNN can't plot long haul conditions if the hole between double cross advances turns out to be excessively huge. LSTM, which can plot long haul conditions by characterizing every memory cell with a set of doors <d, where d is the memory measurement of the shrouded territory of LSTM, and it doesn't experience the ill effects of disappearing or detonating angle while performing back spread through time.[3] LSTM contains three entryways, which are elements of xt and ht−1: input door it, overlook entryway ft, and yield entryway ot. The gates together settle on the memory update component. Condition (3) and (2) indicate the measure of data to be disposed of or to be put away from and to store in memory. Condition (5) signifies the yield of the cell ct.

$$it = \sigma (Wi[ht-1, xt] + bi) \quad (2)$$

$$ft = \sigma (Wf [ht-1, xt] + bf) \quad (3)$$

$$qt = \tanh (Wq[ht-1, xt] + bq) \quad (4)$$

$$ot = \sigma (Wo[ht-1, xt] + bo) \quad (5)$$

VII. STRUCTURAL DESIGN (DFD)

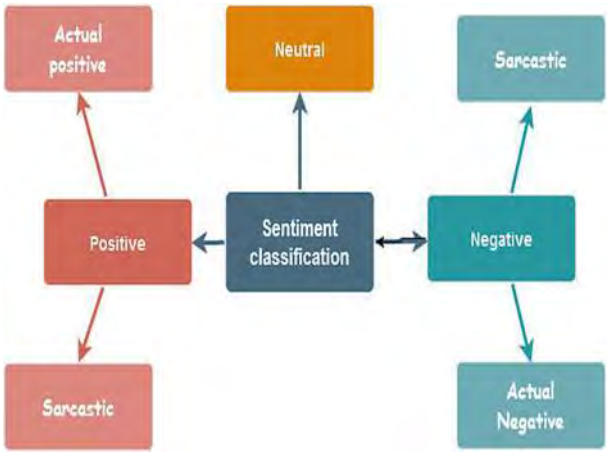


Fig7.1 Structural Design

As from the figure 7.1 as sarcasm location is a subset of sentiment examination. From the outset sentiment examination is ordered into three segments. Each area has its different kind, the three areas are in particular positive, negative, and impartial. In positive the objective information might be genuine positive or sarcastic in negative the objective information might be really negative or it might be sarcastic. Similarly, as with the assistance of sentiment examination the extremity of the objective information is resolved yet because of sarcastic in nature of target information the genuine is turned around extremity.

VIII. EXPECTED OUTCOME:

The yield characterizes the sarcastic and non-sarcastic tweets from the given rundown of tweets. Sentiment Analysis arrange the tweets into positive tweet or negative tweet. The neural models show altogether better exactness contrasted with the comparing discrete baselines. Utilizing likewise setting tweet includes, the exactness of the neural model will be improved, demonstrating the quality of the history data. As in Deep learning the machine naturally extricates the highlights with the assistance of loads of the hubs, inclination, initiation capacity and its ability of forward and in reverse proliferation which encourages it to lessen the cost/mistake. Likewise, with assistance because of our discoveries from different research papers it has been seen that Long-Short-Term-Memory and Gated Recurrent-Neural-Network are promising than Recurrent Neural Network. Profound learning approaches are preferable as far as exactness over regulated learning draws near.

IX. FUTURE SCOPE

Sarcasm Detection in future will be useful to explore and recognize wrongdoing against singular character through advanced methods and stage like twitter and so forth will be useful to forestall against digital bulling .As sentiment analysis is the subset of sarcasm Detection along these lines it will be useful for the online business proprietor and shopper by separating the key terms from the utilization audits which are get posted by client. In future this framework may get computerize and it will be useful to naturally expel mocking tweet without manual intercession.

X. CONCLUSION

Sarcasm is a mind-boggling wonder and it is phonetically and semantically rich. Sarcasm discovery is a captivating State-of-the-craftsmanship. It assesses differing highlight types for sentiment extraction including sentiments, words, examples and n-grams, affirming that each element type adds to the sentiment arrangement structure. Right now, will utilize a Gated Recurrent Neural Network and Long-Short-Term-Memory which gives a superior precision and are intended for the most part message type information. As in Deep Learning dissimilar to Machine Learning, machine removes includes naturally with the assistance of different parameters, for example, loads of the hubs, actuation capacity, and predisposition, forward and in backward propagation to limit the cost/blunder. As identifying sarcasm from content information is a difficult work as it requires to recognize the right feelings of the creator, his history remarks/tweets and furthermore what was nature when such movement happen.

Refrences

[1]. Roger J. Kreuz and Gina M. Caucci. 2007. “Lexical influences on the perception of sarcasm.” In Proceedings of the Workshop on Computational Approaches to Figurative Language.

[2]. Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. “Sarcasm as contrast between a positive sentiment and negative situation”. In EMNLP, pages 704–714.

[3]. “Fracking Sarcasm using Neural Network” Aniruddha Ghosh University College Dublin aniruddha.ghosh@ucdconnect.ie Tony Veale University College Dublin tony.veale@ucd.ieProceedings of NAACL-HLT 2016, pages 161–169, San Diego, California, June 12-17, 2016. c 2016

[4]. “Automatic Sarcasm Detection: A Survey”

[5]. ADITYA JOSHI, IITB-Monash Research Academy PUSHPAK BHATTACHARYYA, Indian Institute of Technology Bombay MARK JCARMAN, Monash University ACM Computing Surveys, Vol.0, No.0, Article 1000.Publicationdate:2017

[6]. “Detecting Sarcasm in Text :An Obvious Solution to a Trivial Problem” Chun-Che Peng CHUNCHE@STANFORD.EDU MohammadLakis MJLAKIS@STANFORD.EDU JanWeiPan JWPAN@STANFORD.EDU Writeup for Stanford CS 229 Machine Learning Final Project. Copyright 2015 by the author(s).

[7]. “An Effective Approach for Detection of Sarcasm in Tweets” Sree lakshmi K Department of Computer Science and Engineering Government Engineering College Palakkad, Kerala, India sreelakshmiknarayanan@gmail.com Rafeeque P C Department of Computer Science and Engineering Government Engineering College Palakkad, Kerala, India rafeeqpc@yahoo.co.in2018 International CET Conference on Control, Communication, and Computing (IC4)

[8]. “Recognition of Sarcasm in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches”

[9]. Piyoros Tungthamthiti, KiyoakiShirai, Masnizah Mohd Japan Advanced Institute of Science and Technology 1-1,Asahidai,NomiCity,Ishikawa,Japan923-1292 Email: {s1320204,kshirai,masnizah}@jaist.ac.jp

[10]. “SARCASM DETECTION IN SENTIMENT ANALYSIS” Shruti Kaushik1, Prof. Mehul P.

Barot2 1 Research Scholar, CE-LDRP-ITR, KSV University Gandhinagar, Gujarat, India 2 Lecturer

[11]. CE-LDRP-ITR, KSV University Gandhinagar, Gujarat, India

[12]. https://medium.com/mlreview/understanding-lstm-and-its-diagrams-37e2f46f1714

[13]. https://commons.wikimedia.org/wiki/File:Gated_Recurrent_Unit,_type_1.svg

[14]. **"**A COMPREHENSIVE STUDY ON SARCASM DETECTION TECHNIQUES IN SENTIMENT ANALYSIS**"** 1Sindhu. C, 2G. Vadivu, 3Mandala Vishal Rao 1,3Department of Computer Science and Engineering, 2Department of Information Technology SRM Institute of Science and Technology Chennai, India 1 sindhucmaa@gmail.com, ISSN: 1314-3395 (on-line version)

Online Web Vulnerability Assessment Scanner

Harshit Sahu, Arushi Saxena, Ankit Sharma, Dr. Sangeeta Vhatkar
Information Technology
Information Technology Thakur College of Engineering and Technology Mumbai, India
harshit16sahu@gmail.com, arushisaxena70@gmail.com, ankit41298@gmail.com, sangeeta.vhatkar@thakureducation.org

Abstract – **The digital transformation of the world is at its peak currently. The internet has had more users in three years than it had ever. This digital transformation has given rise to many online frauds, theft and hacks because of the low security of web applications. As the number of users online has rapidly increased over the years, their privacy and security are a major concern. The loopholes in the security is mostly because of old versions of the software being used and poor coding style. These concerns can be minimized if there is an online tool which constantly monitors the application to find vulnerability using the known threats. There are many precautions that are available in the market but prove to be inefficient because of poor coding of the application. Our aim is to provide an advance penetration testing using the built-in Kali Linux tools through a simple easy-to-use web interface where in the user puts in the URL of the website from the front-end and the rest of the hassle of parsing the URL is handled by Kali Linux server which will be remotely hosted on the internet. The report will be generated and will be stored as a file on the server which can be sent to the user.**

I. INTRODUCTION

The world is experiencing a digital age. Almost everything we do today is digitalized and we are going paperless. Everything is online therefore there is an urgent need to protect those content from hackers. The web application is prone to several attacks which might lead to data loss, data theft, data integrity loss etc. In the current scenario, according to DOS arrest, more than 90 per cent of the websites that are existing on the internet are subjected to various flaws and vulnerabilities. Most of the vulnerabilities are because of the outdated version of the software used for the web application stack. OWASP (Open Web Application Security Project) focuses on the top 10 vulnerabilities that are existing on the web. The project is to detect top 10 OWASP vulnerability in a Web Application. Firstly, the user inputs the URL/IP address of the website. User will be shown several options for the scan that it can perform. The active (attack mode testing) and passive(safe mode) testing can be selected by the user. The front-end dashboard will show and store the result in the database, and the user also receives an email for it. The log reports will be generated and command to test the desired application will be triggered. The actions of the tools are triggered, and the targeted website is scanned for any known vulnerabilities standardized by CVE. We test the application for more than 200+ vulnerabilities that also includes top 10 OWASP vulnerabilities. The various vulnerabilities will be listed down with their potential threat along with their remedy solutions to the user. An email will also be generated in the meantime, which will be sent to the user.

The top 10 vulnerabilities are:

a. Injection
b. Broken Authentication and Session Management
c. Sensitive Data Exposure
d. XML External Entry
e. Broken Access Control
f. Security Misconfiguration
g. Cross-site Scripting
h. Insecure De-serialization
i. Using Components with Known Vulnerabilities
j. Insufficient Logging and Monitoring

We in our project "Vulnerability Assessment Scanner" are

trying to cover all these OWASP vulnerabilities through a simple web interface where in the user simple puts in the target URL in the input box provided. The input box then shoots the POST request on the Linux server which will be intercepted by our NODE JS server running in the backend. The NODE JS server will parse the URL and execute a shell script which will have the commands for running the tests. The Shell Script will accept the URL as a parameter to the function that we write.

II. BACKGROUND

The main objective of this project is to develop a vulnerability assessment scanner that will detect the vulnerabilities present in the web application and generate a report of the same and send it to the user via mail.It is implemented as a system in which a user inserts the URL of the web site, he/she wants to check. The system scans the URL and detects the vulnerabilities present in the system.The system will perform active and passive attacks and based on that; the result will be generated. The system will combine various algorithms that will be required to detect different types of vulnerabilities.

III. RELATED WORK

1. Development

The application will use the following stack for the development.

a. Kali Linux x64 AWS instance hosted remotely
b. React.js front end for user dashboard and login
c. Node.js for handling the backend of the application
d. Php for the Shell Script Calls
e. External JS Libraries

The final deployed application will be a KALI Linux instance running on a cloud server (preferably AWS) that has the frontend and the backend logic written. The front-end of the website will provide the user with a simple interface that allows the user to enter the URL of the website that is to be scanned.

The POST request from the front-end will be proxied and sent to the NODE.js backend which will parse the URL of the user. After parsing the URL, a function will be triggered through the backend of the application. The function will contain all the necessary commands that are required to run the test which are written in a shell script. The tests are for the vulnerabilities which could be found on target URL. On triggering the tests, an asynchronous call would be made to the URL and all the shell script files will start running. The process will be asynchronous, the user on the front end will be shown a loading screen. The user will also be notified by an email once the tests are completed.

The shell script will contain the commands in KALI Linux. After the execution of those commands, an output.txt will be generated which is stored on the server and in NODE JS directory. The results will be generated in a few minutes. The backend NODE on receiving the file will parse it and send the content to the front-end. It also sends an email to the user with the generated report. The user will log in to his dashboard and see the vulnerabilities that are listed.

2. Importance of the project

The project aims at creating a system that scans the vulnerability present in the website or web applications.

In an era where nothing is completely secured, it is unfair to discard the importance of the security. This project aims to make a website free of all the vulnerabilities by performing various active and passive attacks and by checking which attacks are being accepted by the system.

One often faces problems checking various kinds of vulnerabilities together using a single software. Therefore, this project will be built in such a way that it will combine all the codes used to check different types of vulnerabilities in a single system.

3. Economic feasibility

The project relies on open source software, such as OpenVAS, Nexpose Community, Niko, Wireshark, Nessus Professional. The functionality provided by these software's is enough to construct a system that will scan all kind of vulnerabilities. This project is very low cost and thus feasible from an economic standpoint. As such, we do not need any outside funding. The project will be expected to make minimal sales, due to the presence of a variety of other similar tools and the slow adoption of software's in a professional Industry.

4. Ethical Backgrounds

While the software is meant to help users to check their system vulnerabilities, there still is potential for improper usage leading to damage. If the user's system is suffering from various attacks, it might be difficult to detect those attacks. This system is very ethical and legal. It follows all the ethical norms and follows the 3 main aspects of security, I.e. Confidentiality, Integrity and Availability.

First, we will collect all the open source tools onto one virtual machine which will be hosted on AWS. The machine will work as a daemon which will wait for the request to initiate. On receiving the target URL, it will distribute it to all the scripting inputs which are installed on the machine. The process will take around 10-40 mins depending upon the size of the website. The scans will be active or passive depending upon the user. If the user chooses an active scan on the internet, then there will be bombardment of http requests on his server. A passive scan is a quite scan with minimum payload to ensure the website is scanned and the performance should not be affected during testing of the application.

Project planning is essential to managing the scope, schedule and budget of the project. For this, we used tools such as MS Excel, MS PowerPoint, online MS project as well as various modelling tools.

Stakeholder perspective is crucial to the success of this project. Part of the reason is that this system is a highly sensitive field, and even the slightest of errors, which are evidently unavoidable in even the most sophisticated software, can lead to cause error in the result report.

IV. METHODOLOGY

Vulnerability Gathering

Vulnerability gathering is the second process of the advanced penetration testing which consist to

collect the vulnerability of the technology used by the web application. Enumerate the vulnerability from client side to server side. The vulnerability also is done purely manually by using some international common database for vulnerability such as CVE database, CWE database, and NIST database. This vulnerability will show to the penetration tester what the latest vulnerability of the technology are and will be able to locate and exploit them if some exploit framework is available or exploit them manually. Vulnerability gathering is crucial process for the advanced penetration testing because, the vulnerabilities collected in this phase must be specific to the technology that the web application is using. As the database for vulnerability is huge a considerable time have to be consumed on each and every vulnerabilities database. As in the gathering information we get to know the technology used in the web application as an example JavaScript. The advanced penetration establishes some gap analysis between the current version used and the latest version. The advanced penetration testing then identify the vulnerability for the current version either cross-site scripting, security misconfiguration, some flaws in the technology used, DDOS attack possible because of certain aspect of the web.

V. RESULT

Implementation of top 10 OWASP Vulnerability

a. SQL INJECTION

Sql injection attack implies injection of malicious SQL queries to web applications. This attack majorly targets the database of the web application. SQL Injection is performed to test whether any vulnerability is present in the application or not. This is done by adding " ' "or leveraging the URL parameter of the web application. . This will confuse the database and the web application will throw an error because the database doesn't know how to handle the query.

b. BROKEN AUTHENTICATION

Broken Authentication is used to bypass the web application authentication procedure. Here, the attacker mainly uses Brute force attack or dictionary attack. Dictionary attacks use a predefined list of username and password whereas brute force uses all the possible combinations to find the correct username and password.

c. SECURITY MISCONFIGURATION

Security misconfiguration is another threat for web application. Security misconfiguration implies that some parts of the application is not in order which results in the emergence of flaws in the application. Unpatched component, non-used page, default configuration, default location and many more form security misconfiguration. Security misconfiguration may reside either at the client side or the server side of the web application. Some examples of security misconfiguration at client can be improper input sanitization, X-frame options not enable, security protocol not implemented such as HTTPS, SSL or TLS or weak https which have a vulnerability in key exchange (DHE).

d. CROSS-SITE SCRIPTING (XSS)

The cross-site scripting happen with a written script in the input field and reflected on the web application and the second was uploading an exe file (virus) into the web application. Cross-site scripting is an attack against a web application which exploit some vulnerability of it. There are many types of cross-site scripting such as reflected cross-site scripting, DOM cross-site scripting. The result of the cross-site scripting is showing below and demonstrate how cross-site scripting is effective.

e. SENSITIVE DATA EXPOSURE

Sensitive data can be any confidential data such as credit or debit card, business confidential information and many more. If the web application doesn't use secure network protocols, sniffing tools can be used to retrieve the data.

f. XML EXTERNAL ENTITIES

The utilization of the xml processor can cause damage to the web application because xml file can contain malicious code and if the file is uploaded, the attacker can exploit the code in the xml file or exploit some dependency. Modifying the URL from the browser directly allows the attacker to edit any content on the website.

g. BROKEN ACCESS CONTROL

Access control implies security mechanism which is either implemented physically or logically to prevent unauthorized access to network, services, databases and many more. Broken access control is contradictory. It is a major negative impact for the web application. Advanced penetration testing uses sqlmap to bypass the login page of a web application as following:

h. INSECURE DESERIALIZATION

Insecure deserialization is the most recent attack that has been included IN OWASP. It allows the attacker to access directory which is not supposed to be index. Insecure deserialization attack enables discovery of source code and file of html pages that

are not index. Insecure deserialization an attack in which web application de serializes untrusted user input.

i. USING COMPONENT WITH KNOW VULNERABILITY

Web application consists of various components. Some of those components might include serious vulnerabilities which can be easily exploited by the attackers. Examples are WordPress module, windows server and many other components. Vulnerabilities can be exploited by developing some frameworks. Metasploit is an already existing framework which can be used by any attacker or unauthorized user.

j. INSUFFICIENT LOGGING AND MONITORING

Insufficient logging and monitoring is the last attack defined by OWASP. To perform an attack, attackers get into the system and search for as many vulnerabilities as present in the application. By doing this if the attacker is not being detected by the web application or web application firewall or SIEM, then it implies poor or insufficient logging and monitoring of web application.

VI. CONCLUSION

To provide assurance of appropriate level of confidentiality, integrity and availability of information, vulnerability assessments are important mechanisms through which organizations can identify potential security exposures. Routine self-assessments in combination with penetration tests helps us identify security critical areas and provide a good picture of how security is managed and improved. Our tool analyzes and assesses different aspects of any known vulnerability and will generate a detailed report and summary of the vulnerability. The exposure will be fixed through an expert, not through this tool. This tool will be a perfect solution for small scaled web application assessment for vulnerability. A system and method for a vulnerability assessment mechanism that serves to actively scan for vulnerabilities on a continuous basis and interpret the resulting traffic in context of policy is provided. Vulnerability information is presented within an enterprise manager system enabling the user to access vulnerability information, recommended remediation procedures, and associated network traffic. A studio mechanism is used to add scanners to the appropriate policies and control the scope and distribution of scans within the target network.

The network security policy information is merged with the output of the vulnerability scanner. Using the network security policy's information, the output of the scanner is filtered, so that redundant information is removed about applications that are sanctioned by policy. This reduction of information reduces the information overload associated with vulnerability scanning technology, especially when continuous scanning is undertaken.

In addition, the policy monitoring system is provided, through said network security policy, with a description of the intended action of the vulnerability scanner. This permits the monitoring system to report the scanner's operation as a normal usage pattern of the network, eliminating the confusion that can result when a vulnerability scan is monitored.

These enhancements make it possible to run the vulnerability scanner in a continuous scanning mode, with both acceptable monitoring and vulnerability scanner results.

The vulnerability assessment system presents current vulnerability state of the infrastructure. When continuous scanning is used, the vulnerability assessment system provides changes to this state based on new vulnerabilities found or existing vulnerabilities removed. Throughout the process the policy adds context in terms specific to the customer applications allowing prioritized action to be taken to remediate problems. The policy further allows fine grain control of which network hosts are scanned, thus adapting easily to customer operating procedures for how scans may be carried out.

VII. ACKNOWLEDGMENT

We would like to express our special thanks to Dr. Sangeeta Vhatkar to give us this golden opportunity and guide us throughout in completing this project. She has helped us to overcome the obstacles related to the implementation of the project.

References

[1]. Ashikali M. Hasan; Divyakant T. Meva; Anil K. Roy; Jignesh Doshi. (Dec 2017). Perusal of web application security approach. Intelligent Communication and Computational Techniques (ICCT), 2017 International Conference on. doi:10.1109/INTELCCT.2017.8324026

[2]. Bharti Nagpal; Nanhay Singh; Naresh Chauhan; Angel Panesar. (2015). Tool based implementation of SQL injection for penetration testing . International Conference on Computing, Communication & Automation(IEEE Conference Publications), 746 - 749, DOI: 10.1109/CCAA.2015.7148509. Dimitris Mitropoulos; Panagiotis Louridas; Michalis Polychronakis; Angelos D. Keromytis. (2017,).

[3]. Defending Against Web Application Attacks: Approaches, Challenges and Implications. IEEE Transactions on Dependable and Secure Computing (IEEE Early Access Articles), Volume: PP(Issue: 99), 1 - 1, DOI: 10.1109/TDSC.2017.2665620.

[4]. Sugandh Shah; B. M. Mehtre. (2014). An automated approach to Vulnerability Assessment and Penetration

Testing using NetNirikshak 1.0. 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies(IEEE Conference Publications), 707 - 712, DOI: 10.1109/ICACCCT.2014.7019182.

[5]. Jai Narayan Goel, Mohsen Hallaj Asghar, Vivek Kumar, Sudhir Kumar Pandey. (2016). Ensemble Based Approach to Increase Vulnerability. Innovation and Challenges in Cyber Security (ICICCS-INBUSH), 2016 International Conference on. doi:10.1109/ICICCS.2016.7542303

[6]. OWASP. (2016, April 26). Testing for SQL Injection (OTG-INPVAL005). (OWASP Foundation Inc) Retrieved Jun 6, 2017, from https://www.owasp.org/index.php/Testing_for_SQL_Injection _(OTG-INPVAL-005)

[7]. Testing for Cross site scripting. Retrieved May 5, 2017,from https://www.owasp.org/index.php/Testing_for_Cross_site_scri

Accent Recognition System

Tushar Sharma, Shweta Sharma, Rishi Panchal, Pranjali Kasture
Information Technology, Thakur College of Engineering and Technology
Mumbai, India
saytotushar@gmail.com, shwetasha03@gmail.com, panchalrishik@gmail.com, kasturepranjali@gmail.com

Abstract—**Speech recognition systems have been extensively improved over the years. However, accent classification remains a highly challenging task. Accent classification technology can be a great benefit to automatic speech recognition applications, telephony- based service centres, immigration offices and in military operations. The application of convolutional neural networks has been an efficient and effective way to solve the accent recognition problem. The English language is spoken by many around the world, with different accents originating from various geographical locations. Accent identification is an important problem in technology today. Unfortunately, accent identification is suboptimal for certain accents.**

Keywords- ASR- Automatic Speech Recognition
CNN- Convolutional Neural Network
MFCC- Mel Frequency Cepstral Coefficients

I. INTRODUCTION

Speech is one of the most important media of communication between humans. Humans use it to express their opinions as well as their feelings and moods. The adaptation, usage, processing and understanding of human speech by computers can be considered a significant challenge in modern societies. Although many achievements and improvements have been made in the automatic speech recognition (ASR) application area, and more specific with its application in Apple's Siri, Google's Assistant and Amazon's Alexa, the issue of accent recognition seems to be a problem for these programs as they can only understand the American English accent. The above applications can recognize speakers of American English with high accuracy but may fail in recognizing speakers of English with Scottish or Irish accent. The problem seems to be more apparent when the speakers are not native English. We at that point normally resort to a constrained emphasize so as to get the ASR to perceive the discourse accurately, which is unnatural and verification that ASR frameworks can in any case be improved. Since highlight is such a pivotal viewpoint in ASR, we were propelled to assemble a complement grouping AI model which could be utilized as a starter step in the ASR pipeline, permitting it to receive a progressively appropriate discourse acknowledgment model adjusted to the recognized emphasize for better execution. Right now, objective is to build up a profound learning model that can distinguish and group a speaker by their anticipated local language. The contribution to our calculation is an articulation of a word by a speaker. [1]

II. LITERATURE SURVEY

Accent Identification by Combining Deep Neural Networks and Recurrent Neural Networks Trained on Long and Short Term Features, paper by Jiao, Yishan & Tu, Ming & Berisha, Visar & Liss, Julie has the key findings presented an accent identification system by combining DNNs and RNNs trained on long-term and short-term features respectively [4]. Deep Learning Approach to Accent Classification written by Leon MakAn Sheng, Mok Wei Xiong Edmund found out the capabilities of the Deep Neural Networks to classify both native and non-native features [10]. Accent Classification Using Deep Belief Networks written by Rishabh G. Upadhyay had the key findings that Origin of country can be classified [11]. Furthermore MFCC feature acts as human ears.. Deep Voice: Real-time Neural Text-to-Speech written by Sercan Arik, Adam Coate, Andrew Gibiansky, & Jonathan Raiman had the major key findings of the 2-layer CNN struck the best balance between accuracy and model complexity, the Italian accent performed around 5% better on average than what we found in our literature review, the improvements in the Arabic and Japanese accents were less significant, and may have been due to our choice of dataset, the Korean accent still performed poorly, consistent with current findings. [7]

III. PROPOSED SYSTEM DESCRIPTION

The system architecture consists of the processes of collecting the voice signals, preprocessing, feature extraction, classification and output in order to identify the accent of the voice signal. The process of pre-processing was necessary because it could affect the results of the experiments. In every machine learning project the need of using clean data is crucial in order to acquire a good result and output from the system. The pre-processing procedure is divided into two stages. The first stage is to convert each mp3 audio file downloaded from the website to a WAV file of 16 kHz, mono channel and 16-bit. The second stage is to normalise the maximum amplitude of each WAV file to -1.0 dB. The aim of the second stage is to ensure as much possible a universal volume level of each audio file. Figure 16 depicts the diagram of the proposed system

A .Data Set

It consists of speakers who read an elicitation paragraph written in English, which contains common English words and difficult English sounds and sound sequences. The elicitation paragraph consists of English words that cover almost all the sounds of English language and contains as many as possible consonants, vowels and clusters of the standard American English. It is worth to note that each speaker was allowed to read the paragraph for some time and after the recoding was applied.

B. Need of Preprocessing Audio Data

The importance of using clean data in the neural network is high. If clean input data is used, then it is most likely that a clean output can be achieved, in this case meaning a good approximation and accuracy of the model of the convolutional neural network. Usually a satisfactory sampling rate for human voice is at 16 kHz and mono. But the fact of existing unnormalised audio signals is dangerous and it is an alert for the engineer to apply normalisation for the maximum amplitude of the signal. The value of -1.0 dB is the default in Audacity and often it is so in order to leave some headroom for a possible final editing of the audio signal. Fourier analysis is fundamentally a method for expressing a function as a sum of periodic components, and for recovering the function from those components. When both the function and its Fourier transform are replaced with discretized counterparts, it is called the discrete Fourier transform (DFT). [2]

IV. METHODOLOGY USED

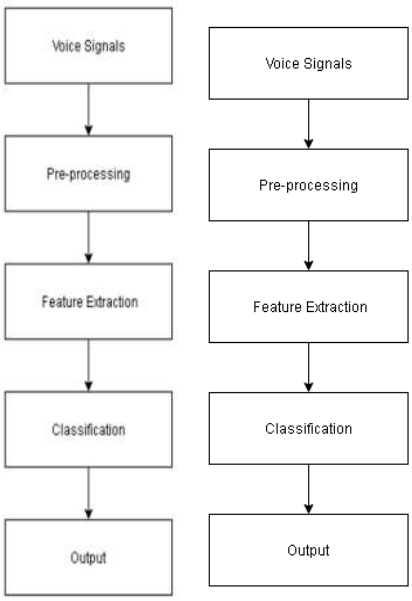


FIG 4.1 ASR FLOWCHART

The pre-processing procedure was a significant process although the dataset from the website was of a good quality. First of all, the downloaded dataset from "The speech accent archive" (Weinberger 2015) was in mp3 format. The proposed system uses audio file of the form of WAV files, sampled at 16 kHz, mono and 16-bit. Winamp was used in order to change the dataset in mp3 format to the desired format. The process is easy as one can load all the audio files in the playlist of the program and set its audio output to Nullsoft Disk Writer and its conversion to PCM 16.000 kHz; 16 Bit: Mono. The second step of the pre-processing was to normalize the maximum amplitude of each WAV files of the speakers. Most of the audio files were quite good concerning their normalization but some of them had low amplitude, which may result in incompetent usage of inputs for the neural network. The program used for the normalization of the audio files was Audacity 2.1.1. And the normalization was also not difficult to implement as one can load all the audio files and apply normalization of the maximum amplitude to -1.0 dB using the batch programming mode of the software.

The importance of using clean data in the neural network is high. If clean input data is used, then it is most likely that a clean output can be achieved, in this case meaning a good approximation and accuracy of the model of the convolutional neural network. Usually a satisfactory sampling rate for human voice is at 16 kHz and mono. But the fact of existing unnormalised audio signals is dangerous and it is an alert for the engineer to apply normalization for the maximum amplitude of the signal. The value of -1.0 dB is the default in Audacity and often it is so in order to leave some headroom for a probable final editing of the audio signal

A. MFCC FEATURES

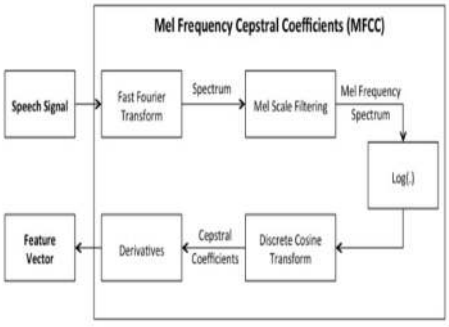


FIG 4.A MFCC ARCHITECTURE

The Mel-Frequency Cepstral Coefficients (MFCCs) of a signal are a small set of features (usually about 10-20) which concisely describe the overall shape of a spectral envelope. In MIR, it is often used to describe timbre. Applications. MFCCs are commonly used as features in speech recognition systems, such as the systems which can automatically recognize numbers spoken into a telephone. MFCCs are also increasingly finding uses in music information retrieval applications such as genre

classification, audio similarity measures, etc. In sound processing, the Mel-Frequency Cepstral (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. Mel-Frequency Cepstral Coefficients (MFCCs) are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip. Mel-Frequency Cepstral Coefficients (MFCC of the real cepstral of a windowed short-time signal derived from the Fast Fourier Transform (FFT) of that signal. The difference from the real cepstral is that a nonlinear frequency scale is used, which approximates the behavior of the auditory system. Additionally, these coefficients are robust and reliable to variations according to speakers and recording conditions. MFCC is an audio feature extraction technique which extracts parameters from the speech similar to ones that are used by humans for hearing speech, while at the same time, deemphasizes all other information. The speech signal is first divided into time frames consisting of an arbitrary number of samples. In most systems overlapping of the frames is used to smooth transition from frame to frame. After the windowing, Fast Fourier Transformation (FFT) is calculated for each frame to extract frequency components of a signal in the time-domain. FFT is used to speed up the processing. The logarithmic Mel-Scaled filter bank is applied to the Fourier transformed frame. This scale is approximately linear up to 1 kHz, and logarithmic at greater frequencies. The relation between frequencies of speech a Mel scale can be established as:

Mel (f) = 1125 ú log (1 + f/700) 2

MFCCs use Mel-scale filter bank where the higher FIG frequency filters have greater bandwidth than the lower frequency filters, but their temporal resolutions are the same. The last step is to calculate Discrete Cosine Transformation (DCT) of the outputs from the filter bank. DCT ranges coefficients according to significance, whereby the 0th coefficient is excluded since it is unreliable. [6]

B. CONVOLUTIONAL NEURAL NETWORK

A CNN architecture is formed by a stack of distinct layers that transform the input volume into an output volume (e.g. holding the class scores) through a differentiable function. A few distinct types of layers are commonly used.

The accent classification problem presented in this thesis was solved by the implementation of two convolutional neural networks. The difference between them is found in the activation functions of their layers.

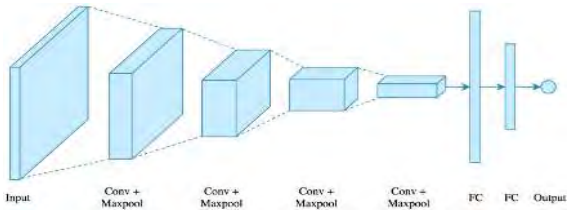


FIG 4.B CNN ARCHITECTURE

Specifically, the first CNN consists of a 2D convolution layer with an output filter of dimension 32 in the convolution, an input shape of 13x30x1 and a ReLU activation function. The data format is set to channels last which means that the ordering of the dimensions in the inputs has the form of (batch, height, width, channels). This includes the number of segmented audio files in batch, the shape of audio files after applying the MFCC with 13 coefficients and the number of channels, which is 1 in the current case. In the domain of image processing this could be seen as an image of dimension 13x30. Next a max pooling operation with a pool size of 2x2 is applied in order to down scale the spatial dimension. The pool size refers to the vertical and horizontal factors of the process of downscaling.

Next a second 2D convolutional layer is applied with the output filter of dimension 64 and the same input shape and activation function as the first convolutional layer. The following step is to apply the same max pooling operation and a dropout operation with a rate of 0.25 in order to avoid overfitting of the network. Next an operation of flattening takes place and the input is flattened into 1 dimension without affecting the batch size. Next a regularly densely-connected layer is added with 128 units and the activation function of ReLU. Another dropout operation is applied next with a 0.5 factor in order to avoid the cases of overfitting. The last layer is a fully connected layer with the number of accent classes used for the model and a softmax activation function.

The second CNN is almost the same as the first with the difference that on the third layer (fully connected) with 128 units instead of ReLU the sigmoid activation function was used. In addition, several other configurations have been tested. These configurations did not result in a better accuracy. However, the two proposed models gave good approximations. [8]

V. RESULTS

If sound frequencies could be turned into images in some manner and sent to eyes to differentiate we might be able to understand a larger range of frequencies. Plotting the audio signal amplitude. The first conversion that can be applied on any sound signal is to plot its samples'

amplitudes over time. Since the dataset contains digits from 0 to 9 we have:

FIG 5.1 PLOTTING A SPECTROGRAM

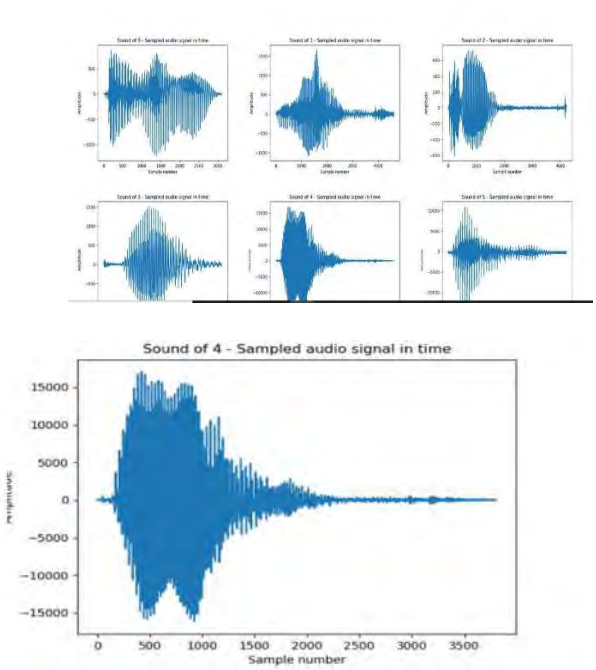


FIG 5.2 AMPLIFIED IMAGE

To plot the spectrogram we break the audio signal into millisecond chunks and compute Short- Time Fourier Transform (STFT) for each chunk. We then plot this time chunk as a colored vertical line in the spectrogram.

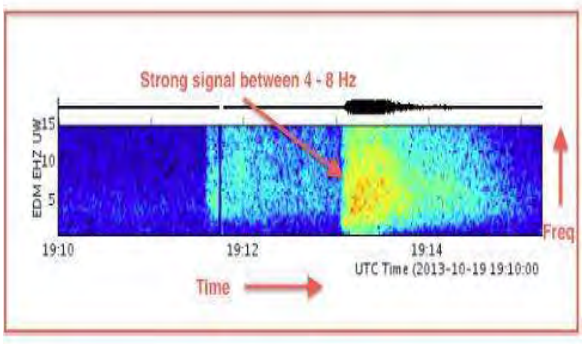


FIG 5.3 TIME CHUNK

Logically both of them can be used to train our CNN. We tried doing that and observed that pure audio signal yields a test-accuracy of 94% as compared to the spectrograms that yield a test- accuracy of 97%.The CNN we use has the following layers

1. Convolution layer with kernel size : 3x3
2. Convolution layer with kernel size : 3x3
3. Max Pooling layer with pool size : 2x2
4. Dropout layer
5. Flattening layer
6. 2 Dense layered Neural Network at the end

VI. CONCLUSION

Speech recognition is the interdisciplinary subfield of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. It is also known as automatic speech recognition, computer speech recognition or speech to text. It incorporates knowledge and research in the linguistics, computer science, and electrical engineering fields. This paper describes an automatic accent identification system.

Our research confirms, this by showing that a DNN system outperforms the GMM system even after an accent- dependent acoustic model was selected using Accent Identification (AID), followed by speaker adaptation. The average performance of the DNN system over all accent groups is maximized when either accent diversity is highest, or data from “difficult” accent-groups is included in the training set.

VII. FUTURE SCOPE

In order to carry out future research, experiments will be performed on the proposed data sets using Deep Belief networks. MFCC features will be employed. Discussion and comparison to these findings would be carried out in order to determine good feature set in combination with new techniques. Features selection will be used to get good results. The results will be compared with various classifiers such as DNN, KNN and ANN. As this classifiers have shown good results in the field of accent classification. Performance will be evaluated on the basis of accuracy of the classifiers. [9]

REFERENCES

[1] Teixeira, Carlos & Trancoso, Isabel & Serralheiro, António. (1996). Accent identification. The 4th International Conference on Spoken Language Processing.

[2] Humphries, J.J. & Woodland, P.C. (1997). Using Accent-Specific Pronunciation Modelling For Improved Large Vocabulary Continuous Speech Recognition. The 5th European Conference on Speech Communication and Technology

[3] T.Chen, C.Huang, E. Chang and J.Wang, "Automatic Accent Identification Using Gaussian Mixture Models", Automatic Speech Recognition and Understanding, 2001. ASRU ’01. IEEE Workshop

[4] Jiao, Yishan & Tu, Ming & Berisha, Visar & Liss, Julie. (2016). Accent Identification by Combining Deep Neural Networks and Recurrent Neural Networks Trained on Long and Short Term Features. INTERSPEECH 2016.

[5] Bower, J.M. & Beeman, D. (1995) The Book of GENESIS: Exploring Realistic Neural Models with the General Neural Simulation System. New York: TELOS/Springer–Verlag.

[6] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. Journal of Neuroscience 15(7):5249-5262.

[7] Sercan Arik, Adam Coate, Andrew Gibiansky, & Jonathan Raiman. (2017) Deep Voice: Real-time Neural Text-to-Speech. ArXiv: 1702.07825

[8] Faria A. (2006) Accent Classification for Speech Recognition. In: Renals S., Bengio S. (eds) Machine Learning for Multimodal Interaction. MLMI 2005. Lecture Notes in Computer Science, vol 3869. Springer, Berlin, Heidelberg.

[9] Alexander, J.A. & Mozer, M.C. (1995) Template- based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), Advances in Neural Information Processing Systems 7, pp. 609– 616. Cambridge, MA: MIT Press.

[10] Leon MakAn, Sheng, Mok Wei and Xiong Edmund- Deep Learning Approach to Accent Classification.

[11] Rishabh G.Upadhyay- Accent Classification Using Deep Belief Networks

Sentiment Analysis On Twitter Data In R

Shubham Maheshwari, Nevil Kaspale, Darshan Faldu mentored by Dr. Anand Khandare
Thakur College of Engineering and Technology, Mumbai, India

Abstract – **Social media has seen a rapid rise since the past decade. The microblogging sector has evolved massively and have emerged out as a vital source of information on one too many occasions. The reason being that these enable people to express their opinions about something or someone in a short manner of expression. Here, in this paper we look into Sentiment Analysis on Twitter Data using text mining and natural language processing with the help of R Programming Language libraries.**

In this paper, we first have a look at what sentiment analysis is, the rise of social media and what are the challenges for researchers. We then write about the related work int his field and discuss about data in Twitter. We then have articulated and visualized our experimentation on performing sentiment analysis on data from twitter. We have used two kinds of datasets

: one is directly extracting real-time current tweets(about a person in our case) and the other is through an external dataset consisting of tweets.

Keywords: Microblogging, Sentiment Analysis, Twitter

I. INTRODUCTION

A. Background

Sentiment Analysis , also referred to as Opinion Mining or Emotion AI, basically means to identify, extract and quantify and study of subjective opinions collected from various sources about a particular subject(s). We know that every person across the planet has some kind of an attitude or 'view' about the things he/she experiences; and since many people experience similar or different kinds of these 'things', there will be dozens of opinions from these billions of people. When all these are viewed collectively, an overall viewpoint about a particular things or product can be ascertained. That was Sentiment Analysis described in general terms.

Rise in interconnectivity through WAN has been of a massive boon to the people. And nowadays, social media platforms like Facebook, Instagram, WhatsApp, Snapchat, YouTube are unignorable[1]. Now, as aforementioned, microblogging websites have seen an exponential rise since the past decade. And it is showing no signs of stopping down as the companies keep optimizing and finding new methods to keep people engaged towards them. Implying more in less amount of words is what is attracting people towards microblogging, where they discuss about, practically anything, be it a public issue, a cricket match, launch of a new kind of technology or an electronic device, and express their view or "sentiment" over it, which can be positive or a complaint (negatory) or maybe neutral[2].

B. Problem Definition

Companies manufacturing these platforms have begun conducting polls in these microblogs to get an idea about user opinions about their product. One challenge for the researchers is to create a technology which gets the company the entire user sentiment data for them and there have been so many successful approaches towards achieving it[2]. We, as well, have taken the most popular microblog, Twitter for our analysis. We have experimented on two kinds of datasets : one being about a person for which we accessed current real time tweets and the second one being an external data set consisting of old tweets and then analysing it. Our results have been visualized for a clearer understanding. We have done all this using R Programming language, as it consists of the necessary libraries with corresponding functions which enable us to extract and analyse Twitter data without much complexity as such.

II. LITERATURE SURVEY

It is vividly evident that the rise of real time information networking sites such as Twitter have led to an assemblage of an unparalleled and a public group of user opinions about every global entity of interest. However, it still poses a challenge as without a tool which analyses these opinions, the user consumption of the product will not expedite. The challenges are mainly attributed to its characteristic of microblogging.

Pak and Paroubek[3] rationale the use of Twitter as a corpus for sentiment analysis. They cited :

People use microblogging platforms to express their opinion and about different topics, thus it is a valuable source of public opinion.

Twitter consists of huge heaps of text and its only growing further. The corpus collected is huge.

Twitter is represented by its users from multiple countries.

Go et. Al[4] proposed distant supervision technique in order to acquire sentiment data. This was initially proposed by Read[10]. They built models using Naïve-Bayes, MaxEnt and Support Vector Machines(SVM) and their training data consisted of tweets with emoticons which served as noisy labels. They report that SVM outperforms other classifiers.

Pak and Paroubek[3] did similar work and classified the tweets as positive, negative
and objective. In order to procure objective data, they retrieved text from Twitter accounts of popular newspapers like "New York Times" and "Washington Post". Their classifier was based on Naïve-Bayes' classifier that uses POS-tags as approach and was based on the N-gram model. Barbosa and Feng[5] were also significant in contributing to sentiment classification. They also classified tweets as objective and subjective and then subjective was further classified into positive and negative tweets. However, they proposed the use of features such as hashtags, retweet, link, punctuation and exclamation marks in conjunction with features such as prior polarity and POS of words.

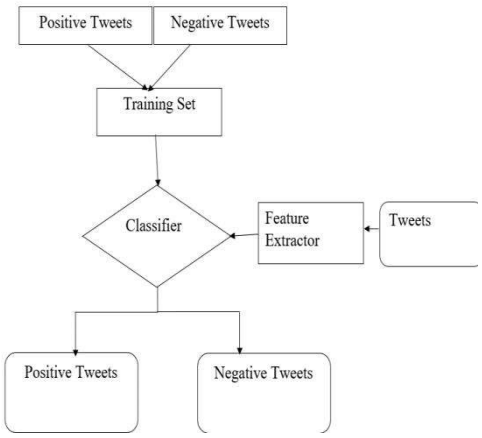
Parikh and Movasatte[6] implemented two Naïve-Bayes unigram models, a Naïve- Bayes bigram model and a Maximum Entropy model to classify tweets. It was found out that the Naïve-Bayes classifier worked better than the Maximum Entropy Model could.

Sentiment Analysis can be done at several levels of granularity, namely, at sentence, phrase, or word level. As twitter permits users to share information in a limited number of characters(280), so it is suited for word level granularity[1].
Further, different approaches have been applied towards automating sentiment analysis in order to predict the sentiments of words, expressions or documents. These include Natural Language Processing (NLP) and Machine Learning (ML) algorithms[2]. In our attempt to mine the sentiment from twitter data here we introduce a hybrid approach which combines the advantages of both dictionary
& corpus based methods along with the combination of NLP & ML based techniques[1].

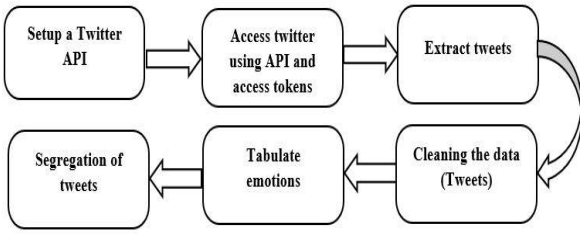
III. DATA DESCRIPTION

Twitter is a social networking and a microblogging platform that allows users topost real time messages, called as 'tweets'. The tweets are now restricted to 280 characters in length. Because of this, people often resort to make spelling mistakes, use acronyms, slangs, emoticons and other characters to express special meaning. Following is a brief terminology with respect to Twitter :
Hashtags: People use "#" to mark topics in order to increase their tweet visibility to a larger audience.
Emoticons: These are representations of facial expressions using punctuations and letters, used to express the mood of the user.
Target : Users use the "@" symbol whenever they are referring to some other user. Using this symbol automatically notifies them.
RT : It is used to indicate that it a user has repeated another user's tweet.

Model Architecture and Data Analysis Opinion words are those which are used by people to express their opinion(positive, negative or neutral). Since our approach is very basic and we have made use of R libraries, our proposed architecture is as :



Our analysis on both datasets works based on the above mentioned architecture. Here, the training set and the classifier(Naïve- Bayes) are all provided by the R library called "syuzhet"[8], developed by Nebraska Literary Lab. It Extracts sentiment and sentiment-derived plot arcs from text using a variety of sentiment dictionaries conveniently packaged for consumption. The text mining is done by the package called "tm".[9]
So our first analysis is on the tweets about our former cricketer Sanjay Manjrekar, who has been gathering a lot of negative limelight lately due to his controversial comments. The flowchart for our process is as follows :



Implementation : Firstly we setup a Twitter API in order to retrieve real-time tweets and then through access tokens we authenticated to the API in R (ROAuth and base64enc packages have their roles here).We went for 50 tweets only under "#SanjayManjrekar". After retrieval, the cleaning of tweets was done so as to remove the unnecessary elements(punctuation, digits, very common words, links etc.). Data cleaning is a very important, although tedious, step towards acquiring sentiments. We then used a function called "get_nrc_sentiments"[8](of syuzhet package) to classify the tweets using string matching into positive, negative and neutral. Then we displayed the tweets that are negative and tweets are positive. Finally, we got the number of positive, negative and neutral tweets, then we visualise them through bar and pie plots. The results are as follows :This is the sample data of the project with keyword "#SanjayManjarekar with 50 most recent tweets. From 50 first six are shown below:
The fetched tweets are classified into three categories: Positive, Negative & Neutral.

```
> head(tweets.df)
[1] "Only dhoni knows how to use Jadeja. He is a very good bowler and a fielder. #sanjaymanjrekar"

[2] " MaheshVasiredd4: imjadeja sanjaymanjrekar Why not send Sanju-Manju to advise Imran how to win elections? He is better than Prashan
t K…"
[3] " thefield_in: #CWC19 \n\nWhy Ravindra Jadeja's Twitter jibe at #SanjayManjrekar is problematic | By vinayakkm \n\n"

[4] " SirJadeja: Don't you dare mess with Sir Jadeja. He'll slay you with his sword and tear you apart into 'Bits and Pieces'. \u2694\U
0001f44f\U0001f44a #SanjayMa…"
[5] "Let this be a learning curve for sanjaymanjrekar how people who have achieved more than him respect each other… "

[6] " GuideforAll: 3rd July 1851 #TodayInHistory\n\nA great social reformer #MahatmaJyotibaPhule founded girls school in \"ChiplunkarWad
a\" #Pune…"
```

fig1:Fetched tweets for #SanjayManjrekar" shown below

The classified tweets are shown below:

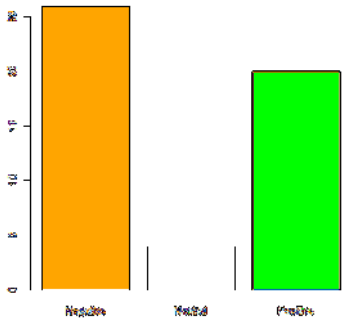
```
> head(positive.tweets)
[1] "Only dhoni knows how to use Jadeja. He is a very good bowler and a fielder. #sanjaymanj
rekar"
[2] " MaheshVasiredd4: imjadeja sanjaymanjrekar Why not send Sanju-Manju to advise Imran how
 to win elections? He is better than Prashant K…"
[3] "Let this be a learning curve for sanjaymanjrekar how people who have achieved more than
 him respect each other… "
[4] " GuideforAll: 3rd July 1851 #TodayInHistory\n\nA great social reformer #MahatmaJyotibaP
hule founded girls school in \"ChiplunkarWada\" #Pune…"
[5] " rohit_hittu: Like and Retweet if you feel this is true.! \U0001f639\n#sanjaymanjrekar
 #jadeja #TeamIndia #CWC2019 #NZvENG #ENGvNZ "
[6] "#sanjaymanjrekar the catches that imjadeja takes you would have not even slipped in you
r bathroom like that so ple… "
```

fig2:Positive tweets

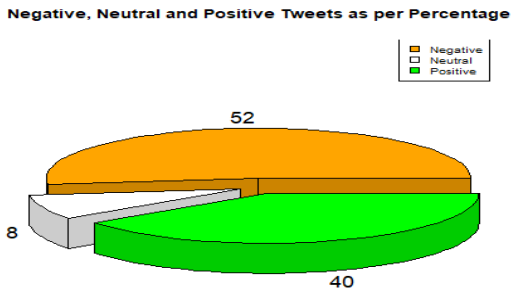
```
> head(negative.tweets)
[1] " thefield_in: #CWC19 \n\nWhy Ravindra Jadeja's Twitter jibe at #SanjayManjrekar is prob
lematic | By vinayakkm \n\n"
[2] " SirJadeja: Don't you dare mess with Sir Jadeja.  He'll slay you with his sword and tea
r you apart into 'Bits and Pieces'. \u2694\U0001f44f\U0001f44a #SanjayMa…"
[3] "Vikram_R_87 sanjaymanjrekar imjadeja Ignore #sanjaymanjrekar. You need not respond to p
etty frustrated mean know… "
[4] " coolsvap: Saifuddin bowls 2 dot balls\n\nSM: Dhoni is struggling\n\nDhoni hits two bou
ndaries\n\nSM: Saifuddin is struggling\n\nI don't know wh…"
[5] " Barry19911: #sanjaymanjrekar How this idiot was even qualified in Indian team \U0001f9
14 Pathetic player! "
[6] " SirJadeja: Don't you dare mess with Sir Jadeja.  He'll slay you with his sword and tea
r you apart into 'Bits and Pieces'. \u2694\U0001f44f\U0001f44a #SanjayMa…"
```

fig3:Negative tweets

The bar chart given below shows the over- all classification of tweets:

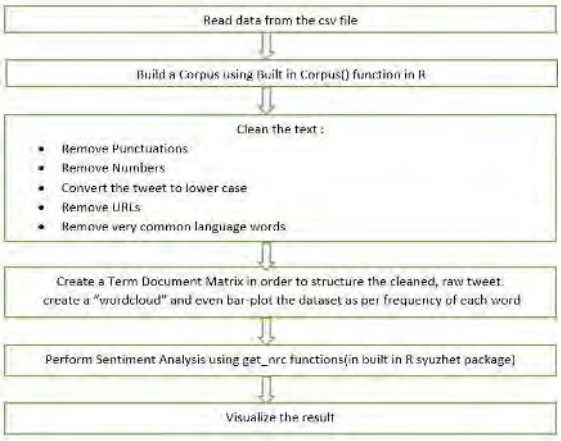


And the pie plot is as :



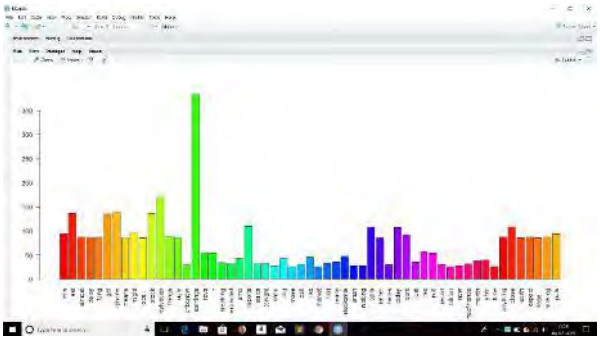

```
library(RColorBrewer) library(wordcloud) library(tm) library(twitteR) library(ROAuth) library(plyr) library(stringr) library(base64enc) library(devtools) library(syuzhet) library(plotrix)
consumer_key - 'mZpdrwGcg4e6hXZCG487JhHIx'
consumer_secret<-
'y8D2dgoMa0fyxA92Z2h9Hiq72cjUlBRynduy9LmZG
QMjLp68
Xi'
access_token <- '763787407040794629-
jjU64asy10NhGDsnRkScCZ9jYYVuRyU'
access_secret<-
'tzbbCVLNkxMbJKZ655aKm2ZnCxVsjeqllDpO4loGX
QnlV' setup_twitter_oauth(consumer_key,
consumer_secret, access_token, access_secret) tweets<-
searchTwitteR('#sanjaymanjrekar',n=50,lang = "en")
n.tweet <- length(tweets)
tweets.df <- twListToDF(tweets)head(tweets.df)
head(tweets.df$text)
tweets.df2<- gsub("http.*","",tweets.df$text) tweets.df2
<- gsub("RT","",tweets.df2) tweets.df2 <-
gsub("https.*","",tweets.df2) tweets.df2 <-
gsub("@","",tweets.df2) head(tweets.df2)
word.df <- as.vector(tweets.df2) emotion.df <-
get_nrc_sentiment(word.df) emotion.df2 <-
cbind(tweets.df2, emotion.df) head(emotion.df2)
sent.value <- get_sentiment(word.df) most.positive <-
word.df[sent.value == max(sent.value)]
most.positive
most.negative <- word.df[sent.value <=
min(sent.value)]
most.negative
positive.tweets <- word.df[sent.value > 0]
head(positive.tweets)
negative.tweets <- word.df[sent.value < 0]
head(negative.tweets)
category_senti <- ifelse(sent.value < 0, "Negative",
ifelse(sent.value > 0, "Positive", "Neutral"))
head(category_senti)
table(category_senti)
a<-table(category_senti)
b<-c('orange','white','green') png(file="BarTw.jpg")
barplot(a,col = b)
dev.off()
x <- c(26, 4, 20)
labels <- c("Negative","Neutral","Positive")
piepercent<- round(100*x/sum(x), 1) png(file =
"city_percentage_legends.jpg") color1 <- c("Orange" ,
"White" , "Green")
pie3D(x, labels = piepercent,explode = 0.1 , main
= "Negative, Neutral and Positive Tweets as per
Percentage",col = c(color1)) legend("topright",
c("Negative","Neutral","Positive"),cex = 0.8,
fill = c(color1))
dev.off()
```

In our second experiment, we took an external dataset of Apple Inc. consisting of
1002 tweets. The data is in the form of csv file. This is a corpus based approach with the use of Natural Language Processing. The Workflow for this is as follows :



Implementation : Firstly, the entire dataset is read from the csv file in R. Then, a corpus (it is just a way to store documents in R readable format). Then text cleaning is done, as explained in the previous example
as well, all unnecessary elements are removed from the tweet. After that, the (cleaned) data which we have is

plain and unstructured, in order to structure it, we need to create a Term Document Matrix. It is a table that contains the frequency of the words. Row names are documents and column are words and. The function *TermDocumentMatrix()* from the text mining(tm) package is used. Bar plot for term document matrix is as :



Then a word cloud[7] is created which helps a lot in text mining as it gets us to know the most relevant words. Wordcloud can be a very handy tool whenever you want to highlight the most commonly cited words. Wordcloud can be of any shape depending on us. Our wordcloud came as :

After changing shape of cloud:

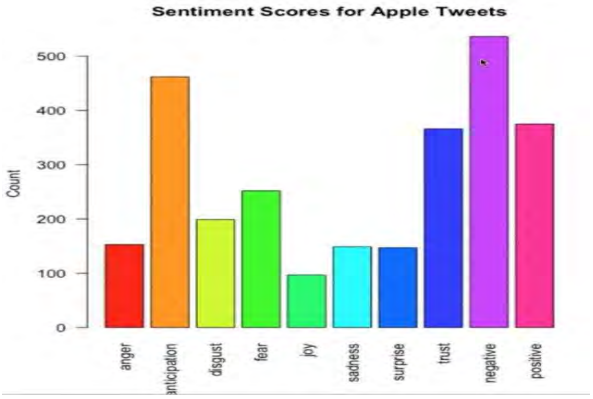
It is evident that 'earnings' is the most relevant word here.Then finally, we get the sentiment scores of the Term Document Matrix (all of the tweets) using the get_nrc_sentiment() function provided by the R library syuzhet. It categorises sentiments into ten categories.
The entire Source Code of R is :

```
#READ DATA
apple <- read.csv(file.choose(), header = T)
str(apple)
#CORPUS
library(tm)
corpus <- iconv(apple$text, to='UTF-8', sub
= "byte")
corpus <- Corpus(VectorSource(corpus))
inspect(corpus[1:5])
#DATA CLEAN
corpus <- tm_map(corpus, tolower)
inspect(corpus[1:5])
corpus <- tm_map(corpus, removePunctuation)
inspect(corpus[1:5])
corpus <- tm_map(corpus, removeNumbers)
inspect(corpus[1:5])
cleanset <- tm_map(corpus, removeWords,
stopwords('english')) inspect(cleanset[1:5]) removeURL
<- function(x)
gsub('http[[:alnum:]]*', '', x)
cleanset <- tm_map(cleanset,
content_transformer(removeURL))
inspect(cleanset[1:5])
cleanset <- tm_map(cleanset, removeWords,
c('aapl', 'apple'))
cleanset <- tm_map(cleanset, gsub, pattern
= 'stocks', replacement = 'stock') cleanset <-
tm_map(cleanset, stripWhitespace)
inspect(cleanset[1:5])

# Term document matrix
tdm <- TermDocumentMatrix(cleanset)
tdm
tdm <- as.matrix(tdm)
tdm[1:10, 1:20]

# Bar plot
w <- rowSums(tdm)
w <- subset(w, w>=25)
barplot(w,
las = 2,
col = rainbow(50))

# Word cloud library(wordcloud)
w <- sort(rowSums(tdm), decreasing = TRUE)
set.seed(222)
wordcloud(words = names(w),
freq = w, max.words = 150, random.order = F, min.freq
= 5,
colors = brewer.pal(8, 'Dark2'), scale = c(5, 0.3),
rot.per = 0.7)

library(wordcloud2)
w <- data.frame(names(w), w)
colnames(w) <- c('word', 'freq')
wordcloud2(w,
size = 0.7,
shape = 'triangle',
rotateRatio = 0.5, minSize = 1)

#letterCloud(w, word = "apple",size=1)
# Sentiment analysis library(syuzhet) library(lubridate)
library(ggplot2) library(scales) library(reshape2)
library(dplyr)
# Read file
apple <- read.csv(file.choose(), header = T)
tweets <- iconv(apple$text, to='UTF-8', sub
= "byte")
# Obtain sentiment scores
s <- get_nrc_sentiment(tweets)
head(s) tweets[4] get_nrc_sentiment('delay')


# Bar plot barplot(colSums(s),
```

las = 2,
col = rainbow(10),
ylab = 'Count',
main = 'Sentiment Scores for Apple Tweets')
The sentiment scores looks as:

```
> head(s)
  anger anticipation disgust fear joy sadness surprise
1     0            0       0    0   0       0        0
2     0            0       0    0   0       0        0
3     0            0       0    0   0       0        1
4     1            0       2    2   0       1        0
5     0            0       0    0   0       0        0
6     0            0       0    0   0       0        0
  trust negative positive
1     0        0        1
2     0        0        1
3     0        0        0
4     0        3        0
5     0        0        0
6     0        0        0
```

And the above result, when visualised as a bar plot looked as :



IV. CONCLUSION

Sentiment Analysis is a valuable advancement in technology that allows us to get realistic feedback from users in a less- biased manner. The work presented in this paper specifies the results for sentiment analysis on Twitter data using the libraries provided by R. The exponential growth in microblogs like this provides the best opportunity to mine sentiments using new technical models and postulates. We investigated two kinds of datasets: first about a person where real time tweets were extracted through API and second was an external csv file consisting of tweets. The corpus-based method was used(in the latter case) to find semantic orientation of adjectives. Then the overall tweet sentiment was calculated and then the result was visualized as well. The domain is exploratory in nature and for future approach, we will do feature analysis and explore on even richer linguistic analysis, as in, parsing, topic modelling and sentiment analysis.

References

[1] Akshi Kumar, Teeja Mary Sebastian, "Sentiment Analysis on Twitter", IJSCI- International Journal of Computer Science Issues, 03 July 2012.

[2] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, "Sentiment Analysis of Twitter Data"

[3] A. Pak and P. Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320–1326.

[4] A. Go, R. Bhayani, L.Huang. "Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper ,2009

[5] Luciano Barbosa and Junlan Feng.2010. Robust sentiment detection on twitter from biased and noisy data. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages36–44.

[6] R. Parikh and M. Movassate, "Sentiment Analysis of UserGenerated Twitter Updates using Various Classification Techniques", CS224N Final Report, 2009

[7] Rohit Nair, *Analysis of Twitter Data Using R Parts 1 through 3,* Medium.com: https://medium.com/@rohitnair_94843/analysis-of-twitter-data-using-r-part-2-word- cloud-dd423af1b2c6, May 2016

[8] Matthew Jockers, *Package "syuzhet",* retrieved from : https://cran.r-project.org/web/packages/syuzhet/syuzhet. pdf, 14 December 2017.

[9] Ingo Feinerer, Kurt Hornik, *Package "tm",* retrieved from : https://cran.r- project.org/web/packages/tm/tm.pdf, 21December 2017.

[10] J. Read. "Using emoticons to reduce dependency in machine learning techniques for sentiment classification". In Proceedings of ACL-05, 43nd Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2005

Smart Garage System

Mr. Akshay Chauhan, Mr. Jemin Nasit, Mr. Shivam Pandey.
Thakur College of Engineering and Technology, Mumbai-India.
tcet@thakureducation.org

Abstract – **The Internet of things (IoT) is the extension of Internet connectivity into physical devices and everyday objects. IoT is a rapidly expanding field in the world of technology. It helps ease life for a person by means of internetworking physical devices, automotive, home appliances and other items embedded with electronics, software, sensors and actuators.In normal garages it is very tedious and challenging task for people to use the pulley or to open the shutter. It not only consumes time but require a lot of human effort. So to overcome with this type of issues we came up with an idea where we can simplify the process and make it feasible for the people. This improvised ideaof accessing your garage will not just save your time but also human effort.Smart garage is one such device that helps ease the access of the garage doors in domestic use. Unlike most solutions, smart garage helps you to access your doors with the help of a computable device which is as simple as a RFID tag to authorize the access. We designed one such garage using Node MCU and a RFID sensor. Which we compared with existing methods. We have illustrated the construction process of the prototype of our garage and the various components required for the same.**

I. INTRODUCTION:

The Internet of things (IoT) is the extension of Internet connectivity into physical devices and everyday objects. IoT is a rapidly expanding field in the world of technology. It helps ease life for a person by means of internetworking physical devices, automotive, home appliances and other items embedded with electronics, software, sensors and actuators. In normal garages it is very tedious and challenging task for people to use the pulley or to open the shutter. It not only consumes time but require a lot of human effort. So to overcome with this type of issues we came up with an idea where we can simplify the process and make it feasible for the people. This improvised idea of accessing your garage will not just save your time but also human effort. Smart garage is

one such device that helps ease the access of the garage doors in domestic use. Unlike most solutions, smart garage helps you to access your doors with the help of a computable device which is as simple as a RFID tag to authorize the access. We designed one such garage using Node MCU and a RFID sensor. Which we compared with existing methods. We have illustrated the construction process of the prototype of our garage and the various components required for the same.

II. BACKGROUND:

A garage door opener is a motorized device that opens and closes garage doors controlled by switches on the garage wall. Most also include a handheld radio remote control carried by the owner, which can be used to open and close the door from a short distance.

The first garage door opener remote controls were simple and consisted of a simple transmitter (the remote) and receiver which controlled the opener mechanism. The transmitter would transmit on a designated frequency; the receiver would listenfor the radio signal, then open or close the garage, depending on the door position. The basic concept of this can be traced back to World War II. This type of system was used to detonate remote bombs. While novel at the time, the technology ran its course when garage door openers became popular. While the garage door remote control transmitter is low power and has limited range, its signal can be received by other, nearby, garage door openers. When two neighbors had garage door openers, then opening one garage door might open the neighbor's garage door as well.

In normal garages it is very tedious and challenging task for people to use the pulley or to open the shutter. It not only consumes time but require a lot of human effort. So to overcome with this type of issues we came up with an idea where we can simplify the process and make it feasible for the people. This improvised idea of accessing your garage will not just save your time but also human effort.

III. IMPORTANCE OF THE PROJECT:

The requirement of the Smart Garage System is because every time it is not feasible to get out of the car to manually open the garage door and close it when the car is parked. Hence there is a great importance of the project in day-to-day life. Importance of Smart Garage System are as below:

- Convenience

One of the main benefits is you can stay in your car until you're in (or out of) your garage, so you don't have to get out of your car to open or close your garage.

- Automatic closing

Automatic closing means that you won't need to get out of your car to close the door or remember to press the button on your remote; your garage door opener can close itself too.

- Low energy use

IV. SUMMARY:

Smart Garage System is boon which helps to overcome the problem of manually opening and closing of garage door while parking the car or removing it from the garage. . IoT is a rapidly expanding field in the world of technology. It helps ease life for a person by means of internetworking physical devices, automotive, home appliances and other items embedded with electronics,

software, sensors and actuators.

The project focus on building an IoT based Smart Garage System using Arduino as platform. Before the project starts there is the first chapter which speaks about the background of the project, why it is required, and it adds about the need for the project which is the removal of manually opening of garage doors. Thus, the project was a great success with all the use cases and fulfilling all the phases of a project development and deployment.

V. LITERATURE SURVEY:

In this paper, Sanjay, B. S. et. al., have expressed the process that the consumer can access the garage just by passing the RFID tag across the sensor. This helps overcome the manual labor of the consumer to access the door. In case of multiple user, a unique tag can be given to any number of authorized person, rather than sharing a selected number keys to the lock or eases to the possibility of losing the key. Replacing PC with low-cost single chip processor can make the cost effectively

Although automatic garage door openers can be highly advanced and technical, they do not need a lot of power to work. This makes them highly cost effective to run, and less damaging to the environment.

VI. OBJECTIVES AND SCOPE OF THE PROJECT:

This project aims to automate the garage door because many times it is not feasible for car owner to get out of the car and open the door manually and close it manually whereas this project helps it to overcome this problem of opening the door manually by making it connect through smart devices.

It can also be used in parking lots where daily tons cars are parked and removed and with help of this project problem of manually opening the door , parking the car and manually closing the door is overcome because with this project the garage door will be automatically gets opened when an authenticated car enters the passage of the garage.

less. Moreover, the Smart garage is accessible to the authorized person, as it is authenticated with a unique RFID tag. This method is more secure than the standard biometric identification method, which can be implemented to the system and hence overcome the problem of security. The Smart Garage also has an IR sensor which detects the presence of the car inside the garage and accordingly notifies the user.[1]

The simplification of the system aims to provide a security access to the restricted and most protected parts of the industry through a system that manages the security of the arena. Margulies, Jonathan mentioned in their paper [2] that several automation models and smart door management systems were considered in modeling the system with the required inclusion relevant for the industrial arena.

Several automation systems for secured access have been introduced in this era of smart phones. Initially several door lock mechanisms were introduced. Further adding to these words Prarthana, R. Jenifer, et al in their paper [3] expresses it is a combo of RFID for maintaining the authentication data, LCD and motor part for maintaining door motions, sensors for inquiring the environment, an inter communiqué module, and control module for establishing overall control

.In this paper, Sunny, R. Jenifer, expresses that garage door is the largest moving object in the home. These doors are often operated by electric door openers. Proper installation, operation, maintenance and testing of the garage door and automatic opener are necessary to provide safe, trouble free operation.[4]

VII. PROBLEM DEFINITION:

The history of the garage door could date back to 450 BC when chariots were stored in gatehouse. The manual method of accessing doors of a garage uses a pulley and a lever or a switch which trigger a motor. The doors are secured with a lock, this arise the difficulty of manually controlling. A fingerprint recognition system is also installed for security purposes preventing unauthorized users

VIII. PHASES OF PROJECT:

Phase 1: Planning, Analysis, Designing and Implementation.

- Analysis: Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation.

- Planning: After analysis we will first study about it and do some research on it for our better understanding of the project and getting a rough picture about what would be our problem definition for the project.

- Designing: Then we will construct the design of the project and according to that, will list down all the requirements needed for the construction for

1. Features of the project:

The Project consists of the following features:

High sensitivity and fast response:

The RFID Tag and module are highly sensitive when the RFID tag comes near the contact with the RFID module and if the Tag is authenticated then the garage door will be opened and if it is not authenticated then the door won't get open.

Simple drive circuit:

The connections and the circuit design of this system is simple hence it is much more convenient to develop.

Long life and stable:

The components are short lived because of the from gaining entry. People may have difficulty accessing the

entry door to their house because it is difficult for them to turn a key or they cannot maneuver a wheelchair into a position to either open the door or grasp the doorknob or scan their finger.

We will use various components such as RFID sensor and the reader for detecting the valid user, Node MCU as a processor which turns out to be the best and most suitable option, Servo Motor for opening and shutting the doors, IR Sensor to detect the presence and absence of the car and used for the internet connectivity. Overall the system provides a complete solution for the problem.

the prototype of our project.

• Implementation: After acquiring the requirements we first developed the connections of the hardware components and then added the coding part later with help of USB cable to the Node MCU.

Phase 2: Testing and Deployment

• Testing: After the prototype is ready we will first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding to it and will check again.

• Deployment: After, complete integration and testing of project real time running and operation of the system will be done. User need to check the value.

components used are of low grade here in this the sensor and the alert system gives alert if the car enters the garage or exit's the garage.

IX. METHODOLOGY:

The proposed method takes input from the RFID Tag and Reader authenticates the tag whether the tag is valid or not. If the tag is valid the door of garage will automatically get opened. Once the car is parked the owner of the car gets the message in mobile. The presence of the car is detected using Infrared Sensor (IR Sensor) . If the car enters the garage then IR has detection for the car hence the owner of the car gets message that the car has entered in the garage and if the car exits the garage then IR has no detection for the presence of the car hence the owner of the car gets message that the car exited the garage.

2. Block Diagram:

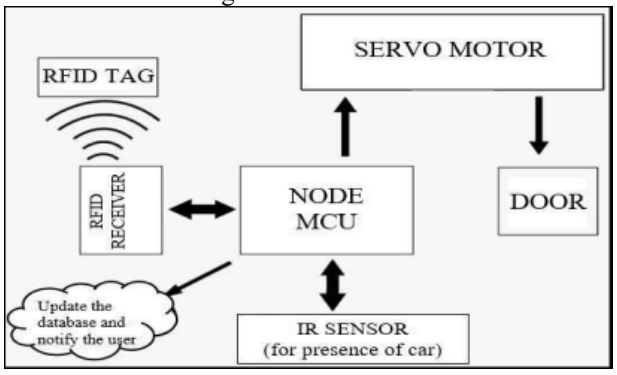


Fig 1: Block Diagram

This Figure shows block diagram of Smart Garage

X. CIRCUIT DIAGRAM:

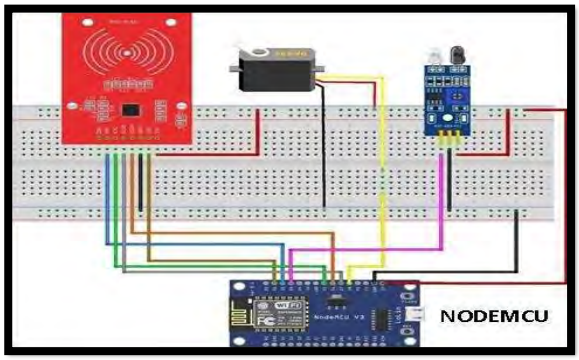


Fig 2: Circuit Diagram

This Figure shows the connection between

System which consist of different Components such as Node MCU, Servo Motor, IR Sensor, RFID tag & Receiver.

different Components such as Node MCU, IR Sensor, RFID Reader and Servo Motor.

XI. SCREENSHOTS:

Fig 3: Smart Garage System

Fig 4: RFID sensor authenticates RFID tag and garage door opens

XII. RESULTS:

Outputs:

- When the car enters the garage it gives notification in application that "CAR ENTERED"
- When the car exits the garage it gives notification in application that "CAR EXITED

Fig 6: Notification in the Blynk application when car enters and exits the garage

XIII. OUTCOMES:

Phase 1:

- Analysis:

Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase.

- Planning:

Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done.

- Designing:

Proposing the design architecture for the prototype.

- Implementation:

Implementation of Node MCU model with the help of code running in the Arduino IDE environment.

Phase 2:

- Testing:

Testing was done with the cases when car wasn't present the garage and what happens when the car enters the pathway of the garage. The testing of the model was a success which is shown in above figures.

- Deployment:

Real time observation was noted when the car entered the pathway of the garage and with the help of RFID reader and tag the car is authenticated and if the car is authenticated the door of garage will be opened automatically.

XIV. IMPROVEMENTS IN RESEARCH GAPS (FINDINGS):

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. They are as follows:

As this system is currently used for automatic opening of door this this project can be further extended to provide features like setting up an alarm which will ring if the garage door is broke openedThe lights inside the garage can also be automated so that whenever the car enters the lights will go on and whenever the car is not present the lights will remain off. Increase the security level of the garage by implementing image processing to read the number plate of the car and process it to check whether the car is authorized to enter inside or not.

XV. FUTURE SCOPE:

The functionality of this project can be further extended to provide features like setting up an alarm which will ring if the garage door is broke opened. The lights inside the garage can also be automated so that whenever the car enters the lights will go on and whenever the car is not present the lights will remain off. Moreover, we can further increase the security level of the garage by implementing image processing to read the number plate of the car and process it to check whether the car is authorized to enter inside or not. This concept of smart garage can be extended to convert many traditional garages to highly secure smart garages.

XVI. CONCLUSION:

We presented Smart garage, an application of loT,

which is used to access the shutter of a domestic garage, with the help of a simple RFID tag, without compromising on the security. As we compare it with the existing solution we can say that, smart garage is effective in terms of mitigating the process and securing. But with more advancement in terms of design and security, we can ensure the ease in consumer's utilization.

References

[1]. Sanjay, B. S., N. Kaushik, V. Srinivasan, S. Prabhakar, and Kayalvizhi Jayavel. "Design and implementation of smart garage-An IoT perspective." In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 2712-2716. IEEE, 2017.

[2]. Margulies, Jonathan. "Garage door openers: An internet of things case study." IEEE Security & Privacy 13.4 (2015): 80-83.

[3]. Prarthana, R. Jenifer, et al. "An Automated Garage Door and Security Management System (A dual control system with VPN IoT & Biometric Database)." 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2018.

[4]. Sunny, R. Jenifer"Garage Door System Safely Guidlines". Door & Access Systems Manufacturers Association International. Archived from the original on 2008-12-23.

Smart Floor Cleaner Robot Using Android-App

Mukesh Sahani, Affan Shaikh, Harshal Goyal
TCET, MUMBAI
mukeshsahani92.ms@gmail.com, affanshaikh85.s@gmail.com, harshalgoyal155@gmail.com

Abstract— **In this project we've tried to implement an RC robot using Bluetooth Module HC-05, Arduino and Motor Driver L298N.It's a wireless Bluetooth control floor cleaning machine and can be amazed at the simplicity and effectiveness of the concept. It's basically DC motors wired on a wheeled wooden container with a cleaning solution placed on top and a scrub attached within the bottom through one among the motors. the comb cleans the ground . Anybody can operate this machine easily. Hence it's very useful in hospitals, houses, etc. Bluetooth module is employed for controlling the complete system with help of remote or mobile. there's an application in mobile for Bluetooth connection between the system and mobile. By using Bluetooth module we are able to direct and switch the system because the user needs. It works great and controlled manually supported the user convenience. Categorically makes cleaning easier and merrier while enabling anyone to make something instead of buy. In modern days interior decorations are getting a very important role in our life. Onslaught of floor may be a important one for our health and reduces the person power constraint. Hence our project is incredibly useful in our day to day life.**

Keywords –, – Arduino UNO;Bluetooth Module, Motor Drivers

I. INTRODUCTION

In recent years, robotic cleaners have taken major attention in robotics research because of their effectiveness in assisting humans in floor cleaning applications at homes, hotels, restaurants, offices, hospitals, workshops, warehouses and universities etc This paper deals with the designing and fabrication of Floor cleaning Machine. The aim of this work is to develop and modernized process for cleaning the ground with wet and dry. it's very useful for cleaning the floors. Floor cleaning system is extremely much useful in cleaning floors in hospitals, houses, auditorium, shops, computer centers etc; it's very simple in construction and simple to control. Anybody can operate this machine easily. It incorporates moisture cotton brush, the comb cleans the ground. Hence it's very useful in hospitals, houses, etc. The time taken for cleaning is extremely less and also the cost is additionally very less. Maintenance cost is a smaller amount. Much sort of machines is widely used for this purpose. But they're working under different principles and the cost is also very high. In this paper a very simple drive mechanism and simple to control any persons. the scale of the machine is additionally portable, so we will transfer from one place to other place very easily. the ground cleaning machine is straightforward, modern house holding device; even children can even operate it easily with safety. it's vital one for every and each houses and hospitals etc. Households of today have become smarter and more automated. Home automation delivers convenience and creates longer for people. Domestic robots are entering the homes and people's daily lives, but it's yet a comparatively new and immature market. However, a growth is predicted and also the adoption of domestic robots is evolving. Several robotic cleaners are available on the market but only few ones implement wet cleaning of floors. the aim of this project is to style and implement a Robot Autonomous and Manual via Phone Application. Cleaner Robot is meant to create cleaning process become easier instead of by using manual cleaners. the most objective of this project is to style and implement a cleaning robot prototype by using Arduino uno, ,motordriver,Bluetooth hc-05 and to attain the goal of this project.

II. PROBLEM DEFINITION

As the task of cleaning floors is done manually, but in the age of growing technology things are getting automated and easier. A number of work has been
done to automate the task of cleaning but it has been implemented on a larger scale requiring more equipment's which in turn require more cost.So to automate to work of cleaning on household level this project has been implemented. This floor cleaning robot requires less component and works on a usual power supply which makes it effective and budget friendly.

III. BACKGROUND

This system is designed with the intention to simplify the process of cleaning by providing remote control to floor cleaning robot. The robot will work as

directed by the user by taking inputs from the provided control.

IV. IMPORTANCE

People desire to use things that are efficient and helpful for reducing their tasks. For keeping this in mind there was the main challenge that robot should clean the floor efficiently and accurately. For this purpose we used water and brushes for pure cleaning.Apart from cleaning another important objective is to develop the robot and it must be of low cost and user friendly. So we designed a robot that is not only cost effective but also easy to operate.

V. OBJECTIVES AND SCOPE

Apart from the small technicalities and conceptual challenges faced in this project, there is one above all, which is meant to serve as a breaking point in the world of common user robotics, amateurs, and engineers, and this is enabling a certain robot that operate automatically in all situations and decide that it can do on its own.

The scrubber design should be modified in future because the current design has few problems. Few of those are the motor is not detachable and the high rpm leads to vibration of the whole system. If these features will be modified, this will work well. The development can be made in the field of sensing and to detect as well as move in the direction of dust and thus resulting in better cleaning of floors. Monitoring, self-charging, lighter body weight and to set alarm on/off time manually are the future scope of this paper.

VI. PHASES

Phase 1: Planning, Analysis, Designing and Implementation.
Analysis: Perform literature survey and find all the research gaps Prepare a report or synopsis to set a path that is to be followed.

After understanding the requirements and making the synopsis the development is to be started.

Designing: In development the first step is to plan on how the project is to be done.

The second step is to set the Arduino and make the connection using connecting wires.

After both the steps are completed properly, the actual development is to be started taking help of the plans and designs created.
We have created a code and the code is applied when the USB is connected to the LCD.
Phase 2: Testing and Deployment.
Testing and Deployment: Various test have being carried by trial and error.

The code has been rechecked and the needed corrections have been made. After Testing and approval, apply the proposed system.

VII. METHODOLOGY

A number of software and hardware implementation techniques were accustomed design and develop the system. Fig. 1 shows the diagram of system. We used a 12VDC motor, Bluetooth HC-05,Motor Driver and Arduino to develop our system. The operation of the robotic cleaner goes to be supported retrieving data from an array of inputs that may tell the condition of the ground space round the scrubber.

VIII. HARDWARE AND SOFTWARE REQUIREMENTS

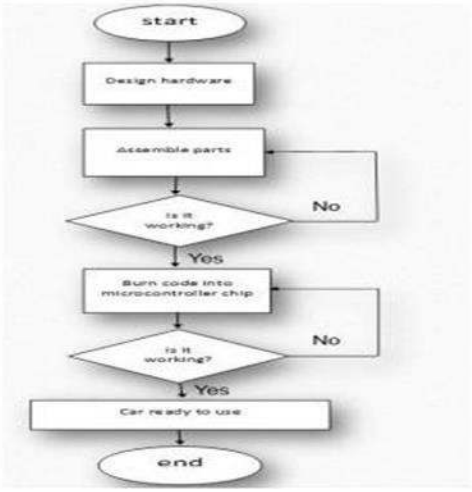
Hardware o Arduino UNO x 1
o Motor Driver L298 Module x 1 o Bluetooth HC-05 Module x 1
o Connecting Wires
o DC Motors x 5

Software

ARDUINO IDE

IX. BLOCK DIAGRAM

Fig1. Block Diagram



The above block diagram represents the actual working of our project. Here the voice input is given to the mic which is processed using the raspberry pi and the reply is given through the speaker in human language.

X. CIRCUIT DIAGRAM

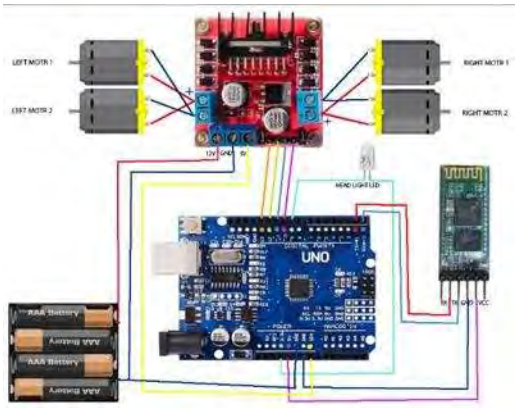


Fig2. Circuit Diagram

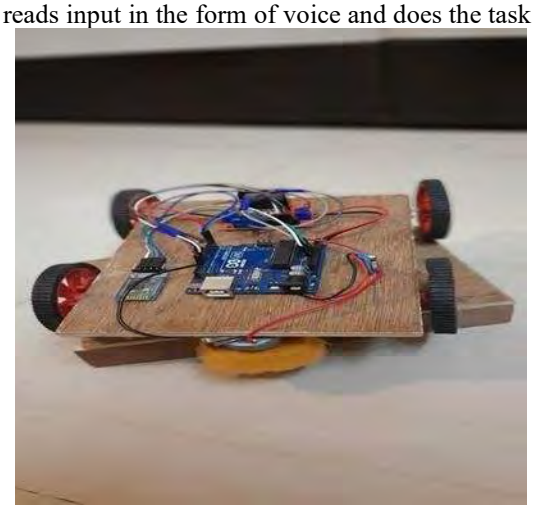
This circuit diagram shows us the connection of different components used in the project. We used Jumper Wires for connecting the components. First we connected the Motor driver with the Arduino
through 10,11,12,13 of Arduino. Then we connected bluettoth module with Arduino. To connect the motors with Motor Driver we did cross connection of
4 motors and connected them .

XI. OUTPUT

When we say OK Google! The assistant starts running and reads input in the form of voice and does the task according to the input received.

XII. RESULT

When we say OK Google! The assistant starts running and reads input in the form of voice and does the task



XIII. OUTCOMES

Creating the outline, works survey and achievability study for proposed solution which will act as an input to the design phase.

Suggesting Statement of work, scope description and scope border for planning the model from the problem definition to decide what needs to be done and what not to be done.

XIV. SUMMARY

Smart floor cleaning system offers a variety of advantages over the traditional cleaning system.The user may sit at a place, start a bot and clean wherever needed. The system consists of a transmitter app. This app is run in an android mobile phone that allows user to transmit command based on user input. Based on these commands the transmitter sends movement commands to the robot. The transmitter is
an android mobile phone that allows user to transmit commands to the robot. On receiving the movement commands from the android. On receiving the movement commands from the android device through Bluetooth receiver. The microcontroller on receiving the commands, decodes them and thus operates the motors in order to achieve the desired motion.This makes floor cleaning a very easy, fast and an effortless process.

XV. CONCLUSION

The project proposed here is an automated android based floor cleaning machine. The system is capable of cleaning the floor cleans using a cleaning mop. The system can work without much loss of human physical energy. The system is provided with an android control which uses Bluetooth communication. The android application can be used to control the robot forward, left, right or back.

This presents that floor cleaning process can be done in an easier manner and more efficiently by robot utilizing wireless robotic system. This proposed robot reduces the time and cost of labor floor.

References:

1. Y.SUNG,R.E.Grinter, and H.I.Chrstensen, and L.Go. " Housewives domestic robot technology", int. Journal of social robotics, 2(4):417-429, 2010
2. Jens-Steffen Gutmann , Kristen Culp , Mario E. Munich and Paolo Pirjanian. The Social Impact of a Systematic Floor Cleaner . In IEEE international workshop on advance robotics an its social impacts ,Technische University munchen, Germany May 21- 23,2012
3. T. Palleja,M.Tresanchez,M. Teixido,J. Palacin" Modeling floor-cleaning coverage performances of some domestic mobile robots in a reduced scenario", Robotics and Autonomous Systems(2010)583745.
4. Xueshan Gao, Kejie Li, Yan Wang, Guangliang Men, Daweizhou and Koki Kikuchi. A floor cleaner robot using Swedish wheels.In IEEE international conference on robotics and biomimetics December 15-18, 2007, Sanya, China
5. J FrolizziC.Disalvo. Service robots in the domestic environment: A study of Roomba
6. Manreet Kaur, PreetiAbrol "Design and Development of Floor Cleaner Robot (Automatic and Manual) "International Journal of Computer Applications (0975 – 8887) Volume 97– No.19, July 2014.

An Introduction of Heartbeat Monitoring System

Mr. Bhavneet Sankhala, Mr. aakash Paliwal, Mr. Rohit Pal,
Information Technology, Thakur College Of Engineering and Technology, Mumbai
sankhalabhavneet1@gmail.com, heyaakashpaliwal08@gmail.com, palrohit03@gmail.com

Abstract— **Technological innovations in the field of disease prevention and maintenance of patient health have enabled the evolution of fields such as monitoring systems. Heart rate may be a very vital health parameter that's directly associated with the soundness of the human circulatory system. Heart rate is that the number of times the guts beats per minute, reflects different physiological conditions like biological workload, stress at work and concentration on tasks, drowsiness and the active state of the autonomic nervous system. It are often measured either by the ECG waveform or by sensing the heart beat - the rhythmic expansion and contraction of an artery as blood is forced through it by the regular contractions of the heart. The pulse are often felt from those areas where the artery is on the brink of the skin. This paper describes a way of measuring the guts rate through a fingertip and Arduino. It is supported the principal of photophelthysmography (PPG) which is non-invasive method of measuring the variation in blood volume in tissue employing a light and detector. While the guts is thrashing, it's actually pumping blood throughout the body, which makes the blood volume inside the finger artery to vary too. This fluctuation of blood are often detected through an optical sensing mechanism placed round the fingertip. The signal are often amplified and is shipped to Arduino with the assistance of interface communication. With the assistance of processing software pulse monitoring and counting is performed.**

Keywords –Arduino; Pulse Sensor; Pulse max; photoelectric plethysmography (PPG)

I. INTRODUCTION

A pulse monitor may be a personal monitor that permits a topic to live their pulse in real time or record their pulse for later study. Early models consisted of a monitoring box with a group of electrode leads that attached to the chest. The heart rate of a healthy adult at rest is around 72 beats per minute (bpm) & Babies at around 120 bpm, while older children have heart rates at around 90 bpm. The heart rate rises gradually during exercises and returns slowly to the remainder value after exercise. The rate when the heart beat returns to normal is a sign of the fitness of the person. Lower than normal heart rates are usually a sign of a condition referred to as bradycardia, while higher is understood as tachycardia. Heart rate is just measured by placing the thumb over the subject's arterial pulsation, and feeling, timing and counting the pulses usually during a 30 playing period . Heart rate (bpm) of the topic is then found by multiplying the obtained number by 2. This method although simple, isn't
accurate and may give errors when the speed is high. More sophisticated methods to live the guts rate utilize electronic techniques. Electro-cardiogram (ECG) is one among frequently used method for measuring the guts rate. But it is an expensive device. Low-cost devices within the sort of wrist watches also are available for the instantaneous measurement of the guts rate. Such devices can give accurate measurements but their cost is typically in more than several hundred dollars, making them uneconomical[1]. So this pulse monitor with a temperature sensor is certainly a useful instrument in knowing the heart beat and therefore the temperature of the topic or the patient.

II. OBJECTIVES:

The goal of this thesis is design low-cost device which measures the guts rate of the topic by clipping sensors on one among the fingers then displaying the result on a android application interface. Miniaturized heart rates monitor system supported a microcontroller. It offers the advantage of portability over tape-based recording systems. The thesis explains how a single-chip microcontroller are often wont to analyze heart beat rate signals in real time. The Hardware and software design are oriented towards a single-chip microcontroller-based system, hence minimizing the dimensions .

III. SCOPE OF THE PROJECT:

The scopes during this project include the hardware and software parts. For the hardware part, ECG circuits are designed so as to interpret data from ECG simulator, which act as a patient. Then, a temperature sensor was developed to live the temperature of person . Both systems are controlled by Arduino Nano board, which connected to the Bluetooth module and wish some programming works. For the software part As and Arduino IDE software are used.

IV. LITERATURE SURVEY

Heart Rate:

The heart is that the organ that liable for pumping blood throughout the body. it's located within the middle of the thorax, slightly offset to the left and surrounded by thelungs basically; the human heart consists of 4 chambers which are two atriums and two ventricles. the proper atrium receives blood returning to the guts from the entire body. That blood passes through the proper ventricle and is pumped to the lungs where it's oxygenated and goes back to the guts through the left atrium of the heart , then the blood passes through the ventricle and is pumped again to be distributed to the whole body through the arteries.[1]

Figure 1.1 heart behavior and part of the generated signal

V. MEASURING THE HEART RATE:

By detecting the R peaks and measuring their frequency, the guts rate are often calculated and then displayed. someone's pulse before, during and after exercise is that the main indicator of their fitness. Measuring this manually requires an individual to prevent the activity they're doing so as to count the amount of heart beats over a period of your time . Measuring the guts rate using an circuit are often done much quicker and more accurately. pulse measurement is one among the vital parameters of the human circulatory system . the guts rate of a healthy adult at rest is around 72 beats per minute (bpm). Athletes normally have lower heart rates than less active people. Babies have a way higher pulse at around 120 bpm, while older children have heart rates at around 90 bpm. the guts rate rises gradually during exercises and returns slowly to the remainder value after exercise. the speed when the heart beat returns to normal is a sign of the fitness of the person. less than normal heart rates are usually a sign of a condition referred to as bradycardia, while above normal heart rates are referred to as tachycardia.[1] Endure athletes often have very low resting heart rates. pulse are often measured by measuring one's pulse. Pulse measurement are often achieved by using specialized medical devices, or by merely pressing one's fingers against an artery (typically on the wrist or the neck). it's generally accepted that taking note of heartbeats employing a stethoscope, a process referred to as auscultation, may be a more accurate method to live the guts rate. There are many other methods to live heart rates like Phonocardiogram1 (PCG), ECG, vital sign wave form and pulse meters but these methods are clinical and expensive.[2] ,[5]

VI. MAXIMUM HEART RATE:

The maximum pulse (Max HR) is that the fastest of heart can beat for one minute. A generalized rule anchors Max HR employing a mathematical formula. Inside each zone, there are different exercise changes which occur because the results of spending training time within the zone. Heart zones is expressed as a percentage of the utmost pulse , reflect exercise intensity and therefore the result benefit.

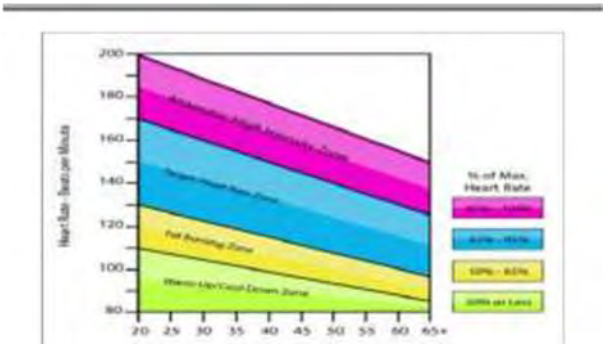


Figure 1.2 Exercise target zone chart

Calculate the utmost rate of the guts is that the subject of ongoing research for an extended time due to the inaccuracy of the calculation. So there are many methods of calculating such

1) The easiest and best known method to calculate your maximum heart rate (MHR) is to use the formula MAXHR = 220 – Age
2) Dr. Martha Gulati et al For Male, MAXHR = 220-Age. For women, MHR= 206 - (0.88 x age).
3) Londeree and Moeschberger • For male and female, MHR = 206.3 - (0.711 × Age) Studies have shown that MHR on a treadmill is consistently 5 to six beats above on a bicycle ergo meter and a couple of to three beats higher on a rowing ergometer. Heart rates while swimming are significantly lower, around 14 bpm, than for treadmill running. Elite endurance athletes and moderately trained individuals will have a MHR 3 or 4 beats slower than a sedentary individual. It was also found that well trained over 50s are likely to have a higher MHR than that which is average for their age.[3],[4]
4) Miller et al For male and female, MHR = 217 - (0.85 x Age)
5) USA Researchers MHR = 206.9 - (0.67 x age)
6) UK Researchers For Male, MHR = 202 - (0.55 x age) For Female, MHR = 216 - (1.09 x age)
7) Miller, Londeree and Moeschberger

To determine your maximum pulse you'll use the subsequent , which mixes the Miller formula with the research from Londeree and Moeschberger.

- Use the Miller formula of MHR=217 - (0.85 × age) to calculate MHR.
- Subtract 3 beats for elite athletes under 30.
- Add 2 beats for 50 year old elite athletes.
- Add 4 beats for 55+ year old elite athletes.
- Use this MHR value for running training.
- Subtract 3 beats for rowing training.
- Subtract 5 beats for bicycle training.

Fingertip sensor:

Use of sunshine to live pulse may be a field of study where abundant research has been wiped out the past few decades. Fingertip sensor relies on measurement of a physiological signal called Photoplethysmogrpahy (PPG) [1,4.], which is an optical measurement of the change in blood volume within the arteries. Fingertip sensor acquires PPG signals by irradiating wavelength of light through the tissue, and compares the light absorption characteristics of blood under these wavelengths.[5]

Photoelectric Photoplethysmogrpahy:

The hardware and software for the MEDAC photoelectric plethysmography (PPG) represent an integrated system for real time monitoring of relative changes in peripheral blood

flow and for recording pulse using a simple to connect sensor. Under appropriate conditions, the software can derive the subsequent measures from the PPG signal: relative blood volume pulse height, pulse wave rise time, pulse wave fall time, the inter-beat- interval (IBI)2, and pulse .

Plethysmography may be a generic term pertaining to a spread of techniques for monitoring volume changes during a limb or tissue segment. Volume changes occur during a pulsatile manner with each beat of the guts as blood flows in and out of some of the body. The study of vascular activity by fluid displacement methods dates back to a minimum of 1890. Photoelectric plethysmography (PPG) was developed in both Germany and therefore the us within the 1930's.

Recent advances in photoelectronics make it possible to utilize photoelectric plethysmography as a sensitive physiological monitoring technique which will be practically applied during a clinical setting.[3]

VII. PROBLEM DEFINITION:

The heart rate monitoring system being developed for the following reasons:

• The goal of this thesis is design low-cost device which measures the guts rate of the topic by clipping sensors on one among the fingers then displaying the result on a android application interface.

• Providing immediate notification of abnormalities in cardiac activity on a monitored patient or user.
• Providing low cost and low power consumption devices provides an inexpensive and reliable method for monitoring patients in developing countries. Providing microcontroller based QRS analysis.
• Providing easily accessible, user friendly and portable device.

VIII. FEASIBILITY:

This project was developed to supply low-cost device which measures the guts rate of the topic by clipping sensors on one among the fingers then displaying the result on a android application interface. This project, as far as feasibility cares , didn't cost an enormous amount because we used Arduino Uno and it's a less price and most of people use for his or her projects. This project are often implemented around a price of 1000 Rs for every setup.

Analysis and Planning

Hardware Requirement:

- **Arduino Uno:**

Fig 1.3 Arduino UNO

Arduino Uno may be a microcontroller board developed by Arduino.cc which is an open-source electronics platform mainly supported AVR microcontroller Atmega328.

The current version of Arduino Uno comes with USB interface, 6 analog input pins, 14 I/O digital ports that are wont to connect with external electronic circuits. Out of 14 I/O ports, 6 pins are often used for PWM output.

It allows the designers to regulate and sense the external electronic devices within the world .

This board comes with all the features required to run the controller and may be directly connected to the pc through USB cable that's wont to transfer the code to the controller using IDE (Integrated Development Environment) software, mainly developed to program Arduino. IDE is equally compatible with Windows, MAC or Linux Systems, however, Windows is preferable to use. Programming languages like C and C++ are utilized in IDE.

Apart from USB, battery or AC to DC adopter also can be wont to power the board.

- **Pulse Sensor:**

Pulse Sensor may be a well-designed plug-and-play heart-rate sensor for Arduino. It are often used by students, artists, athletes, makers, and game & mobile developers who want to simply incorporate live heartrate
data into their projects. The sensor clips onto a fingertip or earlobe and plugs right into Arduino with some jumper cables. It also includes an open-source monitoring app that graphs your pulse in real time.

Fig 1.4 Pulse Sensor

1) A 24-inch Color-Coded Cable, with (male) header connectors. you will find this makes it easy to embed the sensor into your project, and hook up with an Arduino. No soldering is required.
2) An Ear Clip, perfectly sized to the sensor. We searched many places to seek out just the proper clip. It are often hot glued to the rear of the sensor and simply worn on the earlobe.
3) 2 Velcro Dots. These are 'hook' side and also are perfectly sized to the sensor. you will find these Velcro dots very useful if you would like to form a Velcro (or fabric) strap to wrap around a fingertip.
4) Velcro strap to wrap the heart beat Sensor around your finger.
5) 3 Transparent Stickers. These are used on the front of the heart beat Sensor to guard it from oily fingers and sweaty earlobes.
7) The Pulse Sensor has 3 holes round the outside edge which make it easy to stitch it into almost anything.

IX. BREADBOARD:

A breadboard may be a solderless device for temporary prototype with electronics and test circuit designs. Most electronic components in electronic circuits are often interconnected by inserting their leads or terminals into the holes then making connections through wires

where appropriate. The breadboard has strips of metal underneath the board and connect the holes on the highest of the board. The metal strips are laid out as shown below. Note that the highest and bottom rows of holes are connected horizontally and split within the middle while the remaining holes are connected vertically.

Software Requirements:
● Database using MySQL

MySQL is that the hottest Open Source Relational SQL management System. MySQL is one among the simplest RDBMS getting used for developing various web-based software applications. MySQL is developed, marketed and supported by MySQL AB, which may be a Swedish company.

MySQL fully satisfies the ACID requirements for a transaction-safe RDBMS, as follows:

Server-side semaphore variables and locking mechanisms act as traffic managers to assist programs manage their own isolation mechanisms. for instance , MySQL's InnoDB engine uses fine-grained row-level locking for this purpose

Design and Implementation

Circuit Diagram:

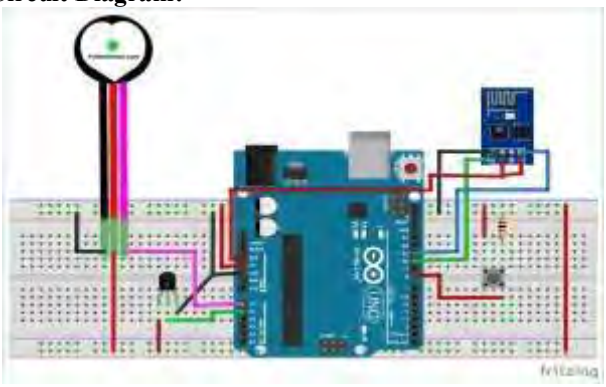


Fig 1.5 Circuit Diagram

X. EXPLANATION:

Circuit of heartbeat monitor is shown below, which contains arduino uno, heart beat sensor module, push button and LCD. Arduino controls whole the method of system like reading pulses form Heart beat sensor module, calculating pulse and sending this data to LCD. we will set the sensitivity of this sensor module by inbuilt potentiometer placed on this module.

Heart beat sensor module's output pin is directly connected to pin 8 of arduino. Vcc and GND are connected to Vcc and GND. A 16x2 LCD is connected with arduino in 4-bit mode.

Flowchart:

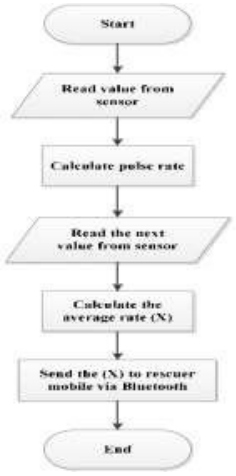


Fig 1.6 Flowchart

The sensor is placed on the hand and measures the heart beat rate when the sensor is activated by the availability voltage. Then, the info is shipped to the controller to look at the heart beat rate through multiple readings. The sensor then transmits the heart beat rate to the swimmer's unit via Bluetooth after half a moment , as displayed in Fig. 1.8

XI. RESULT &DISCUSSION

Result:

The Arduino board is related to the guts beat sensor, presently after finishing the microcontroller and sensor setup the board must be related to an influence source. Since here we utilize sequential correspondence for showing the result or the recognized heart beat we are interfacing the microcontroller with the PC through the USB port. Presently the program is aggregated and transferred into the Arduino board utilizing Arduino compiler and therefore the outcome is accordingly acquired within the sequential screen of the alleged compiler and furthermore the BPM is decided . Presently, the detected information are often put away into an exceed expectations sheet with the help of PLX-DAQ device and that we will change over the straightforward.

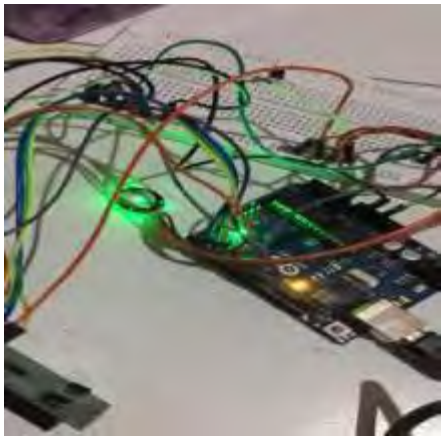


Fig 1.7 working of pulse sensor

Working of Pulse Sensor:

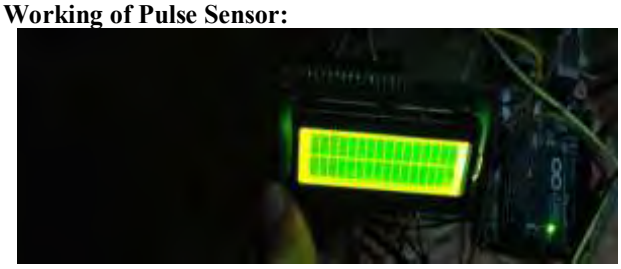


Fig 1.8 LCD displaying BPM

First, we'd like to connect the heart beat Sensor to any organ of body where it can detect the heart beat easily like finger, ears etc. Then the heart beat Sensor will measure the change in volume of blood, which occurs when whenever heart pumps blood within the body. this alteration in volume of blood causes a change within the candlepower through that organ. The Arduino will then convert this alteration into the guts beat per minute (BPM). The LED connected at pin 13 also will blink consistent with the guts Beat.

Conclusion & Future scope

This chapter discusses the suggestion of future work for the project and conclusion are going to be made consistent with the project development. This thesis has discussed the event of the fingertip sensor and interfacing with microcontroller and other peripherals units.

XII. CONCLUSION:

Biomedical engineering (BME) combines the look and problem solving skill of engineering with medical and biological sciences to reinforce patient's health care and so the standard of lifetime of people. disorder is one in all the most causes of untimely deaths in world, heart beat readings are far and away the only real viable diagnostic tool that may promote early detection of cardiac events. By using this we'll measure ones pulse through fingertip.
This paper focuses on the center rate monitoring and alert which is during a position to observe the center beat rate condition of patient. The system determines the center beat rate per minute then sends short message service (SMS) conscious of the mobile . . it's portable and price effective. it is a really efficient system and very easy to handle and thus provides great flexibility and is a superb improvement over other conventional monitoring and alert systems.

XIII. FUTURE WORK:

To ensure the accuracy of pulse rate monitor device, more testing may be performed to larger number of individuals with different ages and weights.

In terms of creating the device more portable, the device would be miniaturized onto a computer circuit board making it light weight and more stable.

We can develop GUI program to point out the center drill PCB. Another alternative we will use free scale microcontroller which use widely in biomedical application to point out the heartbeat because they need GUI program that showing the signals at ADC channels.

the info acquisition system can to become decentralized which suggests localized data acquisition systems that are a part of the PCB of this device can be built that wirelessly transmit the signals collected to the decentralized data acquisition system or on to the processing centre.

References:

[1]. IOT: web of things, https://en.wikipedia.org/wiki/Internet_of_things.

[2]. Compelling Ways to Use Internet of Things in the Field of Medical and Smart Health Care: Kaleem ullah,Munam ali Shah, Sijing zhang.

[3]. IReHMo: An Efficient IoT-Based Remote Health Monitoring System for Smart Regions, Ngo Manh Kohi, Karan Mitra , 2015 seventeenth International Conference on E-wellbeing Networking, Application and Services (HealthCom). Portable based Horne Automation utilizing Internet of Things(IoT), Kumar Mandula, Ram parupalli, 2015 International Conference on Control, lnstrumentation, Communication and Computational Technologies.

[4]. Heartbeat checking and ready framework utilizing GSM innovation, Ufaraoh S.U,Oranugo C.O, International Journal of Engineering Research and General Science Volume 3, Issue 4, July-August.

[5]. Pulse checking and heart assault location, Mamidi manisha, Katakan neeraja, International Journal of Innovations in Engineering and Technology (IJIET)

Dynamic Traffic Management System

M.Abhishek Soni, Ms. Aarushi Singh
Department of Information Technology,
TCET,Mumbai.
abhisheksoni1998@gmail.com, aarushisingh.lm10@gmail.com

***Abstract*—Internet of Things (IoT), is one of the current trends in the Branch of Information Communication Technology. In today's world, urban mobility is one of the unprecedented challenges to be dealt in the administration of a big city. This documentation analyses the ever growing urban population around the globe and discusses about the traffic systems in densely populated cities like Mumbai and Delhi. Further, a traffic management system is proposed and implemented using IoT. The old classical traffic management systems are static i.e just based on predefined structure they operate over a loop. In our Traffic Management System we try to implement a dynamic system that based on current iteration values modify the system over these inputs for next iteration. The main objectives include creation of a Dynamic Traffic management system that will operate according to the number of cars and also detect the presence of vehicles over a range in consideration that the traffic is not steady. After studying the literature survey it was planned to overcome the research gaps and the system was tested using the test cases and then the outputs were generated which justifies that effective time allotment was observed for more congested lanes, preceding over the static system that wasn't capable for the same.**

Keywords – IoT, Dynamic Traffic Management System , Piezo/IR Sensors,Traffic Congestion,Smart City.

I. INTRODUCTION

Traffic signal management is one of the major problematic issues in today's world. In today's scenario, every lane gets a timing of about 1 minute on the road at a regular interval, even when traffic on that particular lane is dense. In our proposed model, we will be optimizing the timing interval of the traffic signal that depends on the number of incoming vehicles on that lane. The major advantage of this system is that it is able to decrease waiting time for the drivers to cross road signal.

One of the important things in the Internet of things in smart cities is the Intelligent Transportation System (ITS).This is possible because of ITS, it utilizes advanced information and communication, and this communication will be helpful for decreasing traffic congestion and to reduce the accidents on the road, which is dangerous in the urban areas.

Managing traffic signal timing is an important thing in urban cities. Managing the time on the road will decrease the waiting time of the drivers on the road, and that will help to reduce the fuel consumption.

A. Importance of the Project

The requirement of the project was to design a dynamic system because in major cities traffic is one of the major problem and our system acts as an alternative to the old classical static system. Also by modifying the green time we would be solving traffic as well as waiting time problems.

Upgrading all the signals in the city will:

- Reduce everyday congestion, by providing time for the lanes that are more congested based on the prior system inputs collected by sensors.
- Reduce pollution throughout the vicinity.
- Developing a dynamic Traffic Management system that fulfills our needs to solve congestion problems and provide smooth journey with less waiting time for the passengers.
- It will provide opportunities to install equipment's/sensors to collect much more detailed traffic information.

You need to carefully consider the benefits that you can achieve by reducing the number of Traffic situations.

B. Motivation

The various issues related to traffic congestion that happened in the previous years. Some of them are as follows:

1. Reaching the office late on the day of an important meeting or discussion, or worse the day of a presentation.

2. Missing a flight due to a traffic jam. This situation has caused people to be hopelessly stuck for many hours, without being able to do anything about it and having to make peace with the fact that they are helpless and can do nothing but sit and wait as the time passes in front of their eyes.

3. Sadly, a traffic jam also causes emergency vehicles like Ambulances or Police Vans to get stuck for a long period of time, which may result in major losses.

4. All these incidents prove that the road infrastructure isn't up to the desired expectations and thus something must be done to deal with these situations.

II. LITERATURE SURVEY

.

A. Literature Survey Table

Paper Title	**Project**	**Author**
Smart Traffic Management System	Article in International Journal of Computer Applications (IJCA) DOI:10.5120/13123-	Ninad Lanke

	0473	
Real Time-Smart Traffic Management System using IoT	Article in International Conference of Emerging Technological Trends(ICETT) DOI: 10.11.09/icett.2016.78736603	Patan Rizwan Kumar Govind Suresh Rajshehara Babu

Few More findings and papers related to the same topic were.

1. Osman et al. recommended a system in which surveillance cameras were used to measure traffic density using MATLAB, a traffic controller and a wireless transmitter used to send images to the server after the server had measured traffic density using those images in each segment. An algorithm was used to set the time span of red light for a specific intersection path, which is calculated by the traffic density on the road and forwarded to the microcontroller and then to the server [1].

2. Jadhav et al. used surveillance cameras, MATLAB and KEIL (Microcontroller coding) to monitor traffic congestion, this paper also addresses priority traffic clearance and red signal broker (Number plate detection). Sensors were used to monitor traffic where vehicle-to-vehicle and vehicle-to-infrastructure contact was conducted using wireless communication systems, controllers positioned at the center of the intersection collected vehicle and pedestrian information and requests and processes using the first-come first-serve method [2].

3. Swathi et al. suggested a smart traffic routing scheme that would choose the shortest route with the least congestion. Sensors are used to collect traffic density data, these sensors use solar energy and batteries, sensors keep transmitting infrared light, and when the object is close, they detect traffic density by tracking the vehicle's reflected light [3].

III. PROPOSED WORK

A. Problem Definition

Traffic congestion leading to more waiting time has been one of the major issues observed by current urban areas since the wide use of vehicles. Only a typical couple of moment's excursion to a store may take up to thirty minutes because of automobile overload or stoppage. As indicated by the police, blockages are really the reasons for certain issues like street rage, street menaces and significant mishaps. The little street limit is additionally one of the contributing components. As the quantity of private vehicles increments incredibly throughout the years, traffic blockage happens when the required street limit isn't satisfied. Improvements of the road can easily solve this problem. Since congestion occurs frequently in the cities, local government municipal can consider passing laws on limiting the quantity of vehicle claimed in a family.

B. Objectives and scope of the project

This project aims for a smarter city with a smarter and better Traffic Management system that will reduce traffic and thus, contribute in the smooth functioning of day to day activities of people living in that city. Our project is a mini replica of the same. The real specimen, however must be tested thoroughly before using it on a large scale in order to test its compatibility. Thus, our project can be expanded on a large scale using better sensors and other advanced forms of technology.

- The project can be further extended by increasing the number of Piezo strips, number of IR sensors at multiple distances to attain more accuracy.
- Also we can determine whether a vehicle has crossed the lane on red light and alert the traffic police with a message for the same.

C. Features of Project

The Project consists of the following features:

1. Dynamicity: Based on input values for different lanes the timings for next iteration are varied accordingly depending on which constraints these values fall.

2. High sensitivity and fast response: The IR sensor and ultrasonic sensor used are highly sensitive to a distant obstacle, along with piezo sensor that takes input on the basis of pressure generated. Output is obtained each point of time in the region as soon as it senses the vehicles nearby.

3. Long life and stable: The system is long lived and much more efficient as compared to the standard traffic management system.

D. Methodology

Our project is divided into two lanes each having different approach to produce the required output by collecting values in two different methods and evaluation for the same using different functions and useful parameters.

For the current iteration, we count the number of vehicles present in Lane 1. For Lane 2, we have IR

readings obtained using the IR sensor, Previous +New-> Ultrasonic Readings. Here, we have used an Ultrasonic

sensor as a rectifier. Based on these values, we decide what amount of green time must be given for the next iteration. At the end of the loop we reconsider all the inputs transform them to useful values and then set all the factors necessary for the change in the system. Now over the next iteration the same is repeated again and again and if the values tend to be minimal no modifications are made in the system.

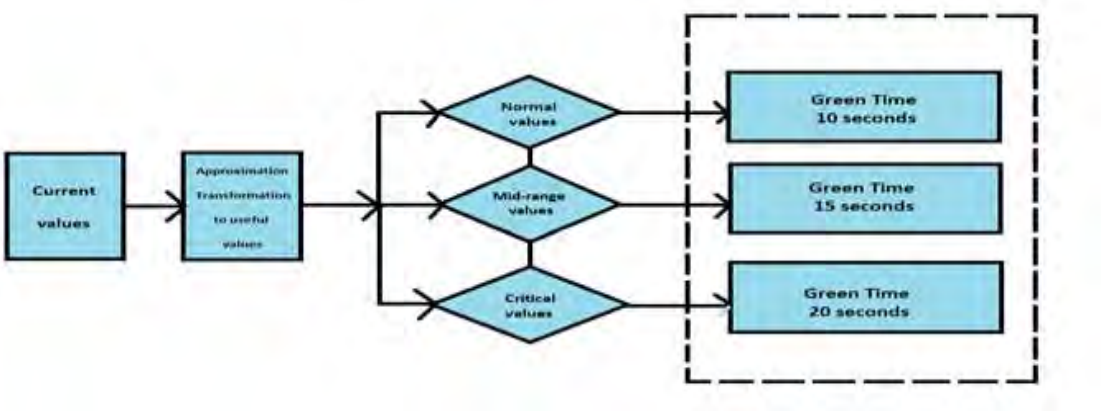


Figure 1: Block Diagram

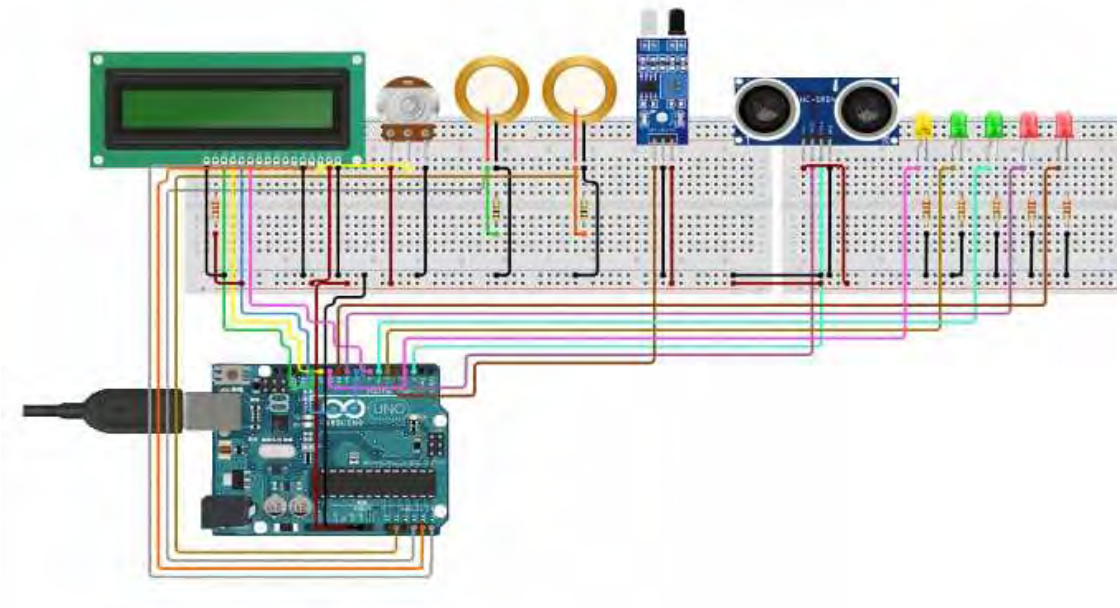


Figure 2: Circuit Diagram

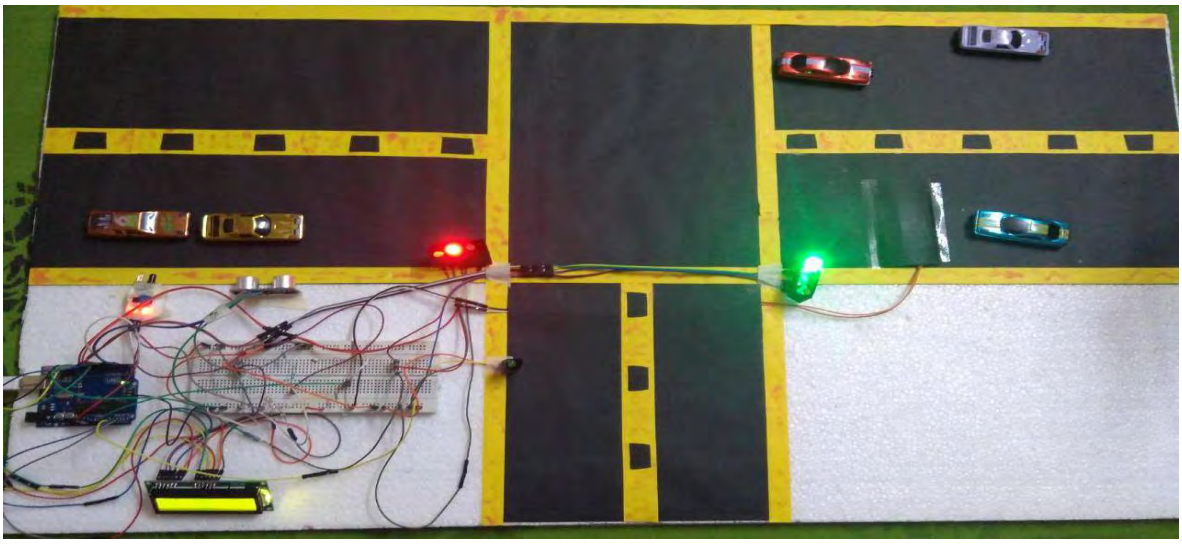


Figure 3: Working Model (Small Scale Prototype)

E. Project Resources(Tools/Components used.)

Hardware Tools

Quantity	Hardware Components
1	Arduino Uno
1	Ultrasonic Sensor
1	IR Obstacle Sensor
2	Piezo Sensor
1	Breadboard, Jumper wires
1	LCD Display (16X2)

Software Tools

1. Arduino IDE 2.Tableau

Along with these additional requirements were soldering of LCD Display, Piezo strip soldering making switches for the traffic signals, etc.

IV. RESULTS AND DISCUSSIONS.

In this project we implemented an Arduino based traffic management model with the help code running on its microprocessor programmed using the Arduino IDE. Testing was done with smaller specimens of vehicles that were passed over piezo sensor or near IR/ Ultrasonic sensor from which outputs were obtained and over those readings Data analysis was done.

Real time sensing was done when the smaller vehicles were passed and the sensor identified the number of iterations correctly and changed the timing of the traffic signal accordingly. Thus the model works efficiently.

Important observations made and results obtained include.

1. This system optimizes the traffic signals over the next iteration based on current inputs to the system.

2. The approximate vehicle count was used to determine green time for the first lane.

3. Simultaneously, a track of the traffic zone in the second lane is kept in consideration that traffic moves over green time and based on that lane 2 timings are optimized.

Figure 4.1: Lane1 Statistical Plot

Figure 4.2: IR Readings and Zone Prediction

V. CONCLUSION

Traffic is one of the major problems faced by the most developed countries. Despite the enormous problem being faced worldwide, almost all the Traffic Management systems are static, i.e. they do not optimize the traffic signals based on current inputs. Our system on the other hand acts as an alternative to this old classical system. This is because on the basis of the current situation of vehicles and traffic zones we have optimized green time, which will be very effective to significantly reduce traffic/ wait time.

This project aims for a smarter city with a smarter and better Traffic Management system that will reduce traffic and thus, contribute in the smooth functioning of day to day activities of people living in that city. Our project is a mini replica of the same. The real specimen, however must be tested thoroughly before using it on a large scale in order to test its compatibility. Thus, our project can be expanded on a large scale using better sensors and other advanced forms of technology.

VI. ACKNOWLEDGEMENT

The author would like to thank his teachers, guides, IT department and everyone who helped him with their guidance, wisdom and support resulting in successful

completion of his project. Also, would like to extend our sincere thanks to all staff in the laboratory for their timely support.

We would like to express our gratitude towards our parents for their kind cooperation and encouragement which helped us in completion of this project.

References

[1] T. e. a. Osman, "Intelligent traffic management system for a cross section of roads using computer vision," in Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual, 2017.

[2] D. C. J. E. J. Khac-Hoai Nam Bui, "Real-Time Traffic Flow Management Based on Inter-Object Communication: a Case Study at Intersection," in 7th International Conference on Ambient Intelligence (ISAmI'2016), 2016.

[3] V. S. A. M. S. D. K. K. Swathi, "Traffic Density Control and Accident Indicator Using WSN," International Journal for Modern Trends in Science and Technology, vol. 2, no. 4, pp. 2455-3778, 2016.

[4] 1. Internet Of Things (A Hands-on-Approach), By Arshdeep Bahga, Vijay Madisetti.

[5] Designing the Internet of Things (Nov 2013) by Adrian McEwen & Hakim Cassimally

[6] "The Internet of Things" by Samuel Greengard Author: Samuel Greengard

[7] www.google.com (Papers , Components, Working, Research)

[8] www.circuito.io (Diagram and basic configuration)

Smart Wardrobe

Khushi Dave, Ami Desai & Radhika Fichadia
Thakur College of Engineering and Technology
khushidave36@gmail.com, desaiami54@gmail.com, fichadiaradhika2@gmail.com

Abstract — **IoT, i.e. Internet of Things, is the currently the most trending field in the Branch of Information Communication Technology. Smart wardrobe is one of its kind implemented with components generally used in automation to enhance functionality of wardrobe. Although similar systems exist for some inventories resembling wardrobe, various features included in this project are original and aren't present in other systems. These features when comes together results in great user experience when it comes to wardrobes which is never seen before. This project is comprised of either freely available software and IDEs for development of required firmware and apps. Hardware installed is widely available and mostly available at a very cheap price. People nowadays like to shopping and buy clothes. However most of the clothes will be storing inside wardrobe for long time even up to several years. Especially in those countries with high humidity, clothes can easily get molded while putting it aside for long time. Smart wardrobe will help them to manage their clothes inside wardrobe. Besides an attached screen, it can also push all data into cloud and further sync up data with mobile application. User can get suggestion on what to wear today from their phone app base on what event are there in their calendar and also the weather. The application will provide a dashboard to show that which cloth having the highest frequency of wearing and which is never touch for long time.**

Keywords – RFID- Radio Frequency Identification

1. INTRODUCTION

Today, individuals have access to an array of smart applications like smartphones, laptops, smart televisions, refrigerators, microwave ovens, etc. However, there aren't many smart devices for the bedroom department of the house. One prominent and important feature that exists in the bedroom is a wardrobe or closet. In fact, over 80% of the respondents we surveyed agreed to have access to a wardrobe. Smart wardrobe helps users manage their clothes inside their wardrobe. Besides having an attached screen to show the status of each cloth, it can also push all data into a cloud and further sync up data with a mobile application. User can get suggestions on what to wear today from their phone app based on the events in their calendar and the weather. In this thesis project, our goal is to put together a set of technologies into a system that could be used to aid the user in the planning of their clothes. In order to achieve this goal, we need to be able to identify which clothes are worn very often and those which are seldom used. We need to categorize our clothes like Casual, Formal, Dinner Party, Business Party, and Social Function and so on. The application will provide statistics to show the clothes that are being worn frequently and those that haven't been touched for a long time. Application can further suggest the user to sell the unwanted clothes or donate them to charity.

2. PROPOSED SYSTEM

To design a microcontroller (IoT) based Smart Wardrobe that will detect the area of clothing present and thereby, will help us locate the clothing materials easily and hassle-free.

A. BACKGROUND

Girls will tend to buy more clothes than their need and ended keep inside wardrobe for years but never wear it. While buying clothes, girls will be controlled by emotion rather than rational. They will likely buy without thinking and this ended up as a result paying more money and occupied more spaces.

Besides, living in the high humidity and warm weather country such as Singapore, new clothes can easily become moldy and smelly.

The idea integrated IoT units together with mobile app. It is not only an inventory tracking system but an established mobile personal inventory management system. With a better management of clothes, user have a better view on which clothes they wear the most frequent at the same time which is least frequent. This may help them to further decide for those least frequent clothes to either give to friend or make a donation for charity. Nowadays RFID technology is becoming more mature and common. The usage of RFID tag has become very common and even printable using a printer. Also, the RFID reader module become cheaper and cheaper at the same time better performance. This can greatly reduce the set up cost of an inventory tracking system. The system only require one time set up and one time registration of cloth, the rest of the process will be automated. Whenever user taking or putting in a cloth, it will be tracked automatically. Once the data is collected, it can be processed further and analyzed to give user suggestion on what to wear today, which cloth has been ignored etc. This concept simply help user understand what their need and help them to save money for buying clothes or suggest them to make a better arrangement for their unused clothes (donate, exchange or sell).

IMPORTANCE

Nowadays RFID technology is becoming more mature and common. The usage of RFID tag has become very common and even printable using a printer. Also, the RFID reader module become cheaper and cheaper at the same time better performance. This can greatly reduce the set up cost of an inventory tracking system.

The system only require one time set up and one time registration of cloth, the rest of the process will be automated. Whenever user taking or putting in a cloth, it will be tracked automatically. Once the data is collected, it can be processed further and analyzed to give user suggestion on what to wear today, which cloth has been ignored etc.

This concept simply help user understand what their need and help them to save money for buying clothes or suggest them to make a better arrangement for their unused clothes (donate, exchange or sell). As a result, small investment bring user great saving.

B. OBJECTIVES AND SCOPE OF THE PROJECT:

Automatic cloth detection:

Using UHF RFID reader instead of LF RFID reader to completely remove the task of user scanning each item manually one by one. This will prevent many fail cases to occur in smart wardrobe which can occur due to causes like scanning multiple items at once.

1. Can fit perfectly in every house and in daily life of people

2. Automatically manage and perform tasks

3. Providing free time for rest of their chores

4. Liberty from repetitive wardrobe management activities

5. Relieving mental stress Affordable.

C. FEATURES OF THE PROJECT

The Project consists of the following features:

1. Simple drive circuit: The connections and the circuit design of this system is simple hence it is much more convenient to develop.

2. Time saving: the project helps us to save time by doing all the search work with the help of RFID tags.

3. Hassle-free: no mess and untidiness to find the clothes which was In the case of traditional method of finding clothes.

3. DESIGN

PHASES OF PROJECT

Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation.

Planning: After analysis we will first study about it and do some research on it for our better understanding of the project and also get a rough picture about what would be our problem definition for the particular project.

Designing: Then we will construct the design of the project and according to that, will list down all the requirements needed for the construction for the prototype of our project.

Implementation: After acquiring the requirements we first developed the connections of the hardware components (Arduino, LEDs etc) and then added the coding part later with help of Arduino data cable to the Arduino.

Phase 2: Testing and Deployment

Testing: After the prototype is ready we will first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding to it and will check again. Deployment: After, complete integration and testing of project real time running and operation of the system will be done. User need to check the value.*Subject Learning Outcomes and Contents*

4. ANALYSIS AND PLANNING

A. Feasibility Study

Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation. After analysis we will first study about it and do some research on it for our better understanding of the project and also getting a rough picture about what would be our problem definition for the particular project.

Technical Feasibility: Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

Economic Feasibility: As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

B. Cost estimation:

Components	Cost
Arduino uno, jumpers, resistors, led,bread board	Rs. 720
RFID-RC522 & RFID tags	Rs. 250
Total	Rs. 970

C. Project planning (Resources, Tools used, etc.)

Hardware Tools:

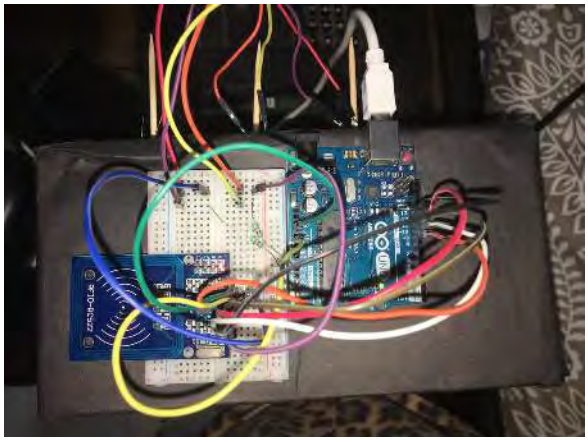
Hardware Components

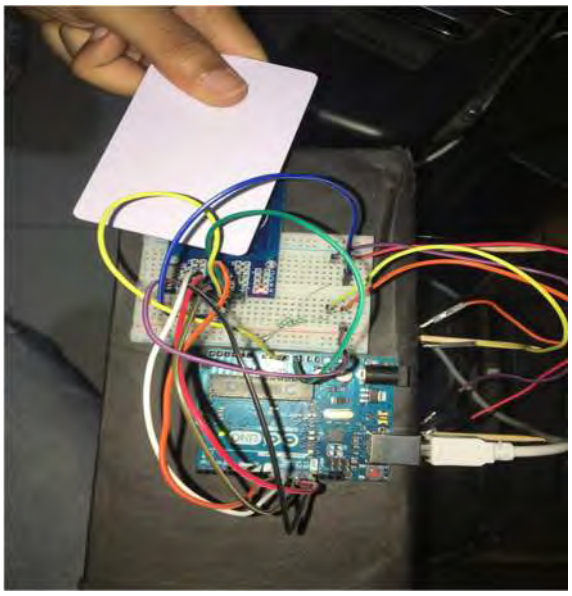
1. Arduino Uno
2. Breadboard
3. Led bulbs
4. Jumpers
5. RFID Reader
6. RIFD Tags
7. Connecting wires

D. Software Tools:

1. Arduino uno
2. RFID Reader

E Result(Expected Ourtcome):





5. CONCLUSION

Thus by this we conclude that the proposed system for smart wardrobe can be used effectively by fashion bloggers, organization etc. Not only an organizer but a normal person can use the system as it is simple to understand. As is, the system may require quite some changes to realize a user experience that would be acceptable but at least we have provided a basis for future research. We have identified challenges and future enhancement.

6. ACKNOWLEDGMENT

We would like to thank subject lecturers (current and previous), education designers and other staff members from the Thakur College of Engineering and Technology who previously designed, developed and taught this subject at various times. The success and final outcome of this project required a lot of guidance and assistance from many people and we am extremely privileged to have completed the project successfully. We would like to thank everyone for their guidance.

References

[1]Adomavicius, G. and Tuzhilin, A. (2001) 'Using data mining methods to build customer profiles', Computer, Vol. 34, No. 2, pp.74–82.

[2]Amato, G. and Straccia, U. (1999) 'User profile modeling and applications to digital libraries', Proceedings 3rd European Conf. Research and Advanced Technology for Digital Libraries, Springer-Verlag, pp.184–197.

[3]Balabanović, M. and Shoham, Y. (1997) 'Fab: contentbased, collaborative recommendation', Communications of the ACM, Vol. 40, No. 3, pp.66–72.

[4]Chen, H.C. and Chen, A. (2001) ' Conference on Information and Knowledge Management,

Predicting Customer Purchase Behavior on Black Friday

Devesh Varma, Harsh Ramani, Angela Roy, Abhishek Yadav
Third Year, Department of Computer Engineering, Thakur College of Engineering and Technology, India

Abstract- **It is no surprise that businesses spend millions of dollars in carrying out market research before coming up with a new service or product. Despite that, it is important to recognize that the final product doesn't sell itself and actually requires the right marketing tools or personalized marketing to make itself visible to the potential consumers. With millions of buying options at our disposal, the behavior of this era's consumers keeps flickering, and it is important to advertise the right product to the right customer at the proper time.**

The project is built with to accurately predict how much will a user spend on a Black Friday sale given his details. The data for our model is a sample of the transactions made in a retail store. We intend to know better the customer purchase behaviour against different products. Specifically, here the problem is a regression problem where we are trying to predict the dependent variable (the amount of purchase) with the help of the information contained in the other variables.

The features of customer like age, marital status, occupation, Current city, will help us accurately predict the amount he/she will spend on a Black Friday Sale. Root mean square Error (RMSE) is used for performance measure. RMSE is commonly used general purpose error metric.

Key Terms- Black Friday, Shopping, Prediction, Multivariate Linear Regression

I. INTRODUCTION

A. Background

The term Black Friday is known to shoppers and retail- ers in the USA as both the day after the Thanksgiving holiday and one of the busiest shopping days of the year the day is characterized by high demand for retail goods. This makes stores to reduce the price on most of its products in order to attract customers due to increased competition. As a result, many stores offer highly pro- moted sales on Black Friday and open very early, such as at midnight, or may even start their sales at some time
on Thanksgiving. Black Friday is not an official holiday, but some states observe "The Day after Thanksgiving" as a holiday for state government employees, some- times in lieu of another federal holiday, such as Colum- bus Day [1].

Black Friday has routinely been the busiest shopping day of the year in the United States since more than a decade [2].
With the use of recent development in Data Science and Machine Learning we can use the details of a person to find his purchase habits, predict his future purchases and use this to give personalized offers to them. This was not possible a few decades ago where the technology was not as advanced as it was today.

B. Problem Definition

Given a data of users such as Gender, Age, City location, Marital Status, Money spent and products bought, build a model to predict the amount of money he will be spending on a Black Friday sale. The main goal is to accurately predict the amount a person spends and the secondary goal is to predict the major factors influenc- ing his spending habits from the given data.

II. LITERATURE SURVEY

A. Machine Learning for Prediction

Machine learning is a subset of artificial intelligence (AI) which learns from experience with respect to some class of tasks, thereby improving its performace with experience. that provides system the ability to automat- ically learn and improve from experience without hav- ing someone programming it. The main goal of machine learning is to focus on the development of computer programs that can access data and use it learn for them- selves. The process of learning begins with observa- tions, facts, information or data, such as examples, di- rect experience, or instruction, in order to look for pat- terns in data and make better decisions in the future based on the examples that are given by the user. The main goal is to allow the computers learn automati- cally without human intervention, help or assistance and adjust actions accordingly. Machine learning with the help of suitable algorithm enables analysis of large quantities of data. Although it generally gives faster and accurate results, it also requires addition resources such as CPU time and memory time to do the necessary pro- cessing and get accurate results [3].

Predictive analytics is a part of Machine Learning which is mainly driven by predictive modelling. It's more of an approach than a process. Predictive analytics and ma- chine learning go hand-in-hand; as predictive models typically include a machine learning algorithm. These models can be trained over

time to respond to new data or values, delivering the results the business needs. Pre- dictive modelling largely overlaps with the field of ma- chine learning.

There are two types of predictive models. They are Classification models, that predict class membership, and Regression models that predict a number. These models are then made up of algorithms. The algorithms perform the statistical analysis, determining trends and patterns in data and predicting the values based upon them. Predictive analytics software solutions have built in algorithms that is used to make predictive models which help in predicting the most accurate value. The algorithms are defined as 'classifiers', identifying which set of categories data belongs to [4].

B. Multivariate Linear Regression Model

The goal of data analysis is to extract meaningful data or information from raw data with accurate estimation. One of the most important and common question con- cerning if there is statistical relationship between a re- sponse variable (Y) and explanatory variables (Xi). An option to answer this question is to employ regression analysis in order to *model* its relationship. Further it can be used to *predict* the response variable for any arbitrary set of explanatory variables.

Multivariate Regression is one of the simplest Machine Learning Algorithm. It comes under the class of Super- vised Learning Algorithms i.e, when we are provided with training dataset [5].

Multiple/Multivariate linear regression models the rela- tionship between two or more known variables and a re- sponse variable by fitting a linear equation to observed data. Every value of the independent variable/varia- bles *x* is associated with a value of the dependent varia- ble *y*.

III. 3. ARCHITECTURE

Stay_In_C	Marital_St	Product_C	Product_C	Product_C	Purchase
2	0	3			8370
2	0	1	6	14	15200
2	0	12			1422
2	0	12	14		1057
4+	0	8			7969
3	0	1	2		15227
2	1	1	8	17	19215
2	1	1	15		15854
2	1	1	16		15686
1	1	8			7871

	User_ID	Product_I	Gender	Age	Occupatio	City_Cate
2	1000001	P00069042	F	0-17	10	A
3	1000001	P00248942	F	0-17	10	A
4	1000001	P00087842	F	0-17	10	A
5	1000001	P00085442	F	0-17	10	A
6	1000002	P00285442	M	55+	16	C
7	1000003	P00193542	M	26-35	15	A
8	1000004	P00184942	M	46-50	7	B
9	1000004	P00346142	M	46-50	7	B
10	1000004	P0097242	M	46-50	7	B
11	1000005	P00274942	M	26-35	20	A

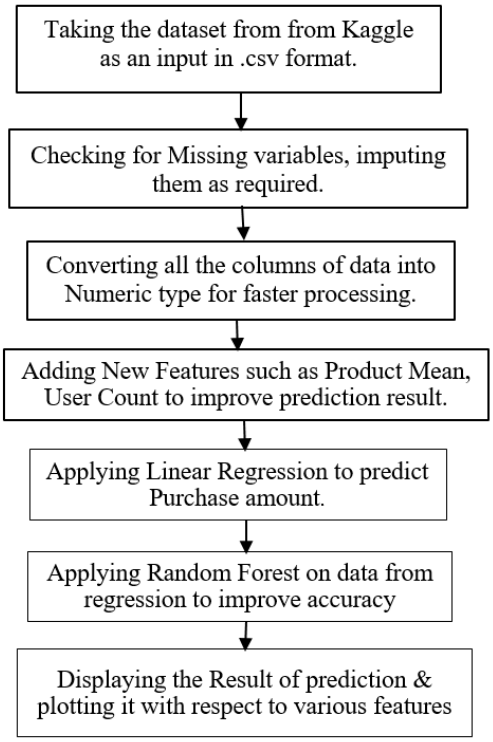
Fig. 1 Structure of Input Dataset

The analysis of the Black Friday data set used in this paper includes 12 variables: Product_ID, Gen- der, Age, City_Category, Stay_In_Cur- rent_City_Years, Marital_Status, Product_Cate- gory_1 and Purchase.

This data set is cleaned from the original data set which contained 3 additional variables - Product_Cate- gory_2 and Product_Category_3 Product_Cate- gory_2 and Product_Category_3 contain missing obser- vations which are imputed as zero.

The Black Friday data set is further cleaned by stating the format of each variable. This included chang- ing Product_ID, Gender, Age, City_Category, Mari- tal_Status and Product_Category from character varia- bles to integers. These variables are nominal since they do not have order or a distance metric. 'Age' is ordinal because the age groups are increasing with actual age, but are not equally spaced.

4. Flow of the Project



IV. IMPLEMENTATION

Source Code:

```
library(caTools)
library(plyr)
library(data.table)
library(DataExplorer)
library(dplyr)
library(ggplot2)
library(vcd)
library(rpart)
library(randomForest)
#Reading Data from CSV file


blackFriday <- read.csv("H:/train1.csv",stringsAsFactors = F)
blackf_data <- fread("H:/train1.csv")

# STEP 1: DATA OVERVIEW head(blackFriday,10)
#What is the structure of the data?
str(blackf_data) # Text
#plot_str(blackf_data) # Network Graph


#STEP 2: DATA PREPROCESSING
 #Filling missing data with zeroes
blackFriday[is.na(blackFriday)]<-0

#Converting all variables to Numeric type
blackFriday$City_Category[blackFriday$City_Category=="A"]<-"1"


blackFriday$City_Category[blackFriday$City_Category=="B"]<-"2"
blackFriday$City_Category[blackFriday$City_Category=="C"]<-"3"
blackFriday$City_Category<-as.integer(blackFriday$City_Category)
# converting age variable to numeric
blackFriday$Age[blackFriday$Age == "0-17"] <- "15"
blackFriday$Age[blackFriday$Age == "18-25"] <- "21"
blackFriday$Age[blackFriday$Age == "26-35"] <- "30"
blackFriday$Age[blackFriday$Age == "36-45"] <- "40"
blackFriday$Age[blackFriday$Age == "46-50"] <- "48"
blackFriday$Age[blackFriday$Age == "51-55"] <- "53"
blackFriday$Age[blackFriday$Age == "55+"] <- "60"
blackFriday$Age <- as.integer(blackFriday$Age)

# converting stay in current city to numeric
blackFriday$Stay_In_Current_City_Years[blackFriday$Stay_In_Current_City_Years == "4+"] <- "4"
blackFriday$Stay_In_Current_City_Years<-as.integer(blackFriday$Stay_In_Current_City_Years)

# converting gender to binary blackFriday$Gender <- ifelse(blackFriday$Gender == "F", 1, 0)
str(blackFriday) blackf_data<-blackFriday #STEP 3: FEATURE ENGINEERING #Adding Product_Mean Feature to our data product_mean <- ddply(blackFriday, .(Product_ID), summarize, Product_Mean=mean(Purchase)) blackFriday <- merge(blackFriday, product_mean, by="Product_ID")

# feature representing the count of each user
user_count <- ddply(blackFriday, .(User_ID), nrow) names(user_count)[2] <- "User_Count" blackFriday <- merge(blackFriday, user_count, by="User_ID")
 #feature representing the count of each product
product_count <- ddply(blackFriday, .(Product_ID), nrow) head(product_count) names(product_count)[2] <- "Product_Count" blackFriday <- merge(blackFriday, product_count, by="Product_ID") #feature representing the proportion of times the user purchases the product more than the product's average
```

```
blackFriday$flag_high <- ifelse(blackFriday$Purchase > blackFriday$Product_Mean,1,0) user_high <- ddply(blackFriday, .(User_ID), summarize, User_High=mean(flag_high)) blackFriday <- merge(blackFriday, user_high, by="User_ID") dt = sort(sample(nrow(blackFriday), nrow(blackFriday)*.8)) train<-blackFriday[dt,] test<-blackFriday[-dt,]

model<-lm(Purchase~Gender+Stay_In_Current_City_Years +Product_Category_2+Product_Category_3 +Product_Mean+User_Count+Product_Count+flag_high+User_High +Occupation,data=train) summary(model) #prediction<-predict(model,test) #blackFriday <- merge(blackFriday, prediction, by="User_ID") model<-randomForest(Purchase~Gender+Age+Marital_Status+Stay_In_Current_City_Years +Product_Category_1+Product_Category_2+Product_Category_3 +Product_Mean+User_Count+Product_Count+flag_high+User_High ,data=train)


plot(model)
pred<-predict(model,test)
class(test)
class(train)
View(test)
View(train)
print(class(pred))
#Tweaking Final Prediction so that the prediction remains in the range
pred[pred<187]<-187
pred[pred>23949]<-23949
final<-cbind(Actual=test$Purchase,Predicted=pred)
as.data.frame(final)->final
View(final)
max(test$Purchase)
min(test$Purchase)
(test$Purchase-final$Predicted)->error
cbind(test,error)->test
rmse<-sqrt(mean(test$error^2))
summary(model)
rmse

blackf_data$User_ID <- as.factor(blackf_data$User_ID)
blackf_data$Product_ID <- as.factor(blackf_data$Product_ID)
blackf_data$Gender <- as.factor(if_else(blackf_data$Gender == 'M', 'Male', 'Female')) blackf_data$Age <- as.factor(blackf_data$Age) blackf_data$Occupation <- as.factor(blackf_data$Occupation)
blackf_data$City_Category<- as.factor(blackf_data$City_Category)
blackf_data$Stay_In_Current_City_Years <- as.factor(blackf_data$Stay_In_Current_City_Years)
blackf_data$Marital_Status <- as.factor(if_else(blackf_data$Marital_Status == 1, 'Married', 'Single')) # Due to the large dataset, the average values were used for this analysis (using pipe operator) new_data <- blackf_data %>% group_by(User_ID, Age, Gender, Occupation, City_Category, Stay_In_Current_City_Years, Marital_Status) %>% summarise_each(funs(mean), Product_Category_1, Product_Category_2, Product_Category_3, Purchase)
# Rename the average values accordingly
colnames(new_data)[8] <- "Product_Cat_1_Avg"
colnames(new_data)[9] <- "Product_Cat_2_Avg"
colnames(new_data)[10] <- "Product_Cat_3_Avg"
colnames(new_data)[11] <- "Avg_Purchase_Amount"

# Age group versus Average purchase amount ggplot(new_data, aes(Age, Avg_Purchase_Amount, fill = Gender)) + geom_col() + facet_wrap(~ Gender) + labs(title = "Age Group/Gender Vs Average Purchase Amount")

# City Category versus Average purchase amount ggplot(new_data, aes(City_Category, Avg_Purchase_Amount, fill = City_Category)) + geom_col() + labs(title = "City Category Vs Average Purchase Amount")

# Stay in current city versus Average purchase amount ggplot(new_data, aes(Stay_In_Current_City_Years, Avg_Purchase_Amount, fill = Stay_In_Current_City_Years)) + geom_col() + labs(title = "Stay in current city Vs Avg_Purchase_Amount")

# Marital status versus Average purchase amount ggplot(new_data, aes(Marital_Status, Avg_Purchase_Amount, fill = Marital_Status)) + geom_col() + labs(title = "Marital Status Vs Avg_Purchase_Amount")
```

V. RESULTS AND DISCUSSION

	Actual	Predicted
7	6187	5959.1365
16	15854	15854.9660
32	5293	4370.1015
34	6914	8411.8718
38	7381	7683.1275
47	19614	14659.7044
48	9872	10107.2901
49	19203	18826.6010
54	11877	14204.1258
56	16662	16446.5617
60	10946	12680.8579
64	12642	15482.4698
66	19681	16202.0383
67	8718	8614.0623

Fig. 2 Actual value v/s Predicted values

The Root Mean Squared Error (RMSE) is:

```
> rmse
[1] 1711.175
```

Gender played a major role in the Amount of Money spent. Men spent three times more when compared to Women.

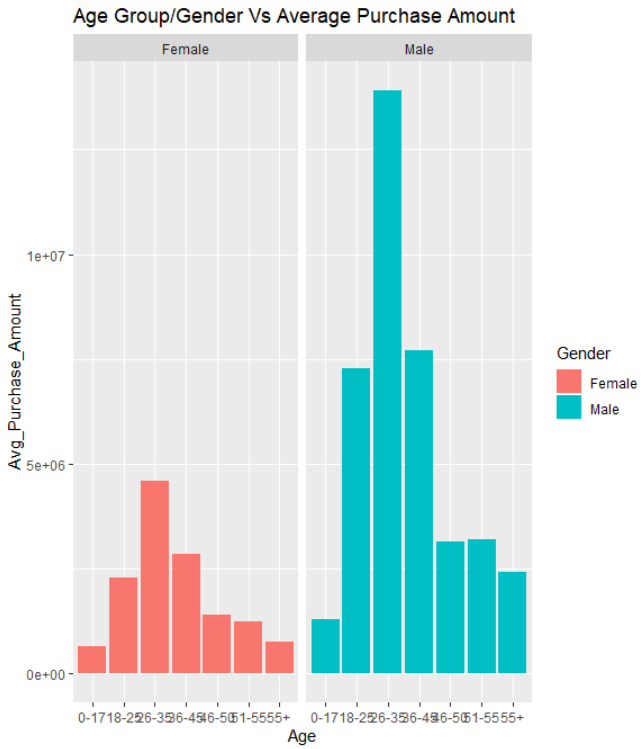


Fig. 3 Graph of Age v/s Purchase Amount

Comparion of which city spent the most money:

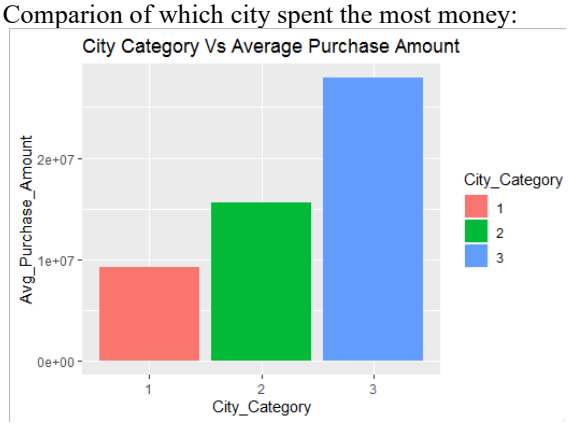


Fig. 4 Graph of City Category v/s Average Purchase Amount

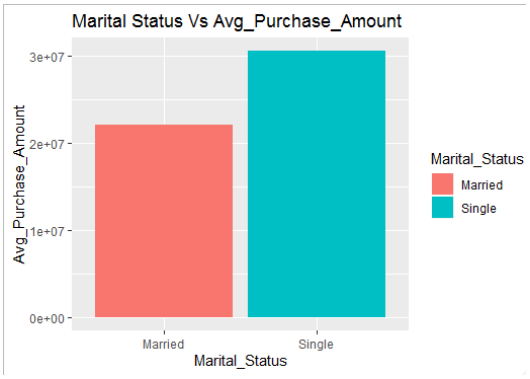


Fig. 5 Relation between Marital Status and Purchase Amount

Thus from our results we have proved some features of a person demographic to decide on how much he spends whereas some features had neglible outcome on the Purchase amount.

References

[1] "Retailers & Sales: Using Social Listening to Learn More About Black Friday – Brandwatch". Brandwatch. December 2, 2014.

[2] International Council of Shopping Centers. "Holi- day Watch: Media Guide 2006 Holiday Facts and Fig- ure". Archived from the original (PDF) on November 29, 2010. ShopperTrak, Press Release,"ShopperTra Reports Positive Response to Early Holiday Promotions Boosts Projections for 2010 Holiday Season". Archived from the original on November 29, 2010. Re- trieved 2012-01-23. (November 16, 2010).

[3] Marco Varone, Daniel Mayer "What is Machine Learning? A definition" http://www.expert- system.com/machine-learning-definition

[4] Katrina Wakefield "Predictive analysis and machine learning" http://www.sas.com/en_gb/insights/arti- cles/analytics/a-guide-to-predictive-analytics-and-ma- chine-learning.html

[5] Abhishek Sharma "Multivariate Regresion" http://www.geeksforgeeks.org/multivariate-regression/

Rapid Remote Development

Parth Ranawat, Aditya Salvi, Amey Sawant, Riddhi Thanki
Information Technology Thakur College Of Engineering And Technology, Mumbai, India
parth.ranawat@gmail.com, adityasalvi3393@outlook.com, ameys2907@gmail.com, riddsthanki@gmail.com

Abstract— **In recent years numerous product development and innovations have been used. From the very beginning, waterfall and V-models have been utilized for programming advancement. But currently there is a noticeable paradigm shift in the way in which software is developed. Hence, we need a product development framework which is rapid and developer friendly. This paper provides a systematic description of Rapid Remote Development.**

Keywords—rapid, remote, development, framework, optimize.

I. INTRODUCTION

Emerging startups are getting progressively mindful of the benefit of offering remote work chances to their employees. Contingent upon the kind of work their employees perform, permitting more remote days or time when employees can work remotely from anywhere other than the place of work, for example, at home or from a local café can profit organizations by expanding efficiency and boosting worker maintenance and productivity. The framework centers around the idea of approved learning and the build-measure-learn feedback loop along with the key idea of remote development. It also attempts to acquire a methodical way to deal with estimating the advancement at a startup. Remote employees are in good company, and as the remote workforce grows, so does the data on its efficacy and the potential benefits for business owners

II. WHY REMOTE DEVELOPMENT

A- Increase in productivity

Several employees say that they are more productive when they work remotely, which could be due, in part, to the fact that many remote workers (58 percent) are able to work from home. All things considered, not every person feels incredible after they work off-site. Albeit over a third of representatives studied appear to feel remote situations are additionally unwinding, 30 percent state they experience expanded pressure while working remotely.

Still, 71 percent say the amount of distraction either stays identical or decreases.

B- Positive impact on workers.

When asked if remote work has a positive or negative impact on various areas in a remote employee's life, answers varied. In general, the capacity to work remotely appears to positively affect most significant regions of a worker's life, including life objectives, family, psychological wellness, and accounts.

Figure 1. Effect of remote development on team members

III. ADVANTAGES

As we mentioned, avoiding the commute to work place is the biggest benefit noted by remote workers (46 percent say this can be the most good thing about remote work).

Having control over their schedules, not having to induce ready for work, and therefore the ability to sleep longer are also benefits. Meanwhile, 28 percent of respondents say having the flexibility to look after their kids could be a good thing about working from home, while just over 25 percent say watching their pets is an important benefit. When received some information about remote work, most of the telecommuters gave it either four or five stars.

1- You're more likely to stay long term

Probably the greatest bit of leeway of remote work is that you're ready to deal with the majority of what life tosses at you—including the moments that might have previously put your position in danger. Going back and forth between work and home feels like a major stressor. Not so with remote work. A laptop means you can work from home and even on the road. Remote occupations additionally imply that in case you're thinking about maternity leave or your partner finds a new line of work in an alternate city, you don't need to forfeit the activity you love for the individual (or future little individual) you love.

2- You'll have more time for deep work

A study came out that proved that work interruptions cost you up to six hours a day. Working from home or a place where you don't have the foggiest idea about the individuals around you implies more clarity of mind. Another reward: remote work is based on the conviction

that you'll complete your work, not that you have to sit for precisely 40 hours every week. The result is that you'll concentrate on quality not amount, which is useful for everybody including the organization.

3- You'll learn some seriously impressive communication skills

One could contend that it's a lot harder to communicate when you work remotely, yet we would really oppose this idea. Working remotely implies that correspondence is everything. It's promptly clear when the manner in which you're communicating isn't working, and you'll be compelled to fix issues much more rapidly. You're going to significantly improve each technique from how you present plans to how you voice disappointments about a collaborator.

4- You'll feel healthier

A recent report from the University of Minnesota found that work environment flexibility brought down pressure and the danger of burnout. So that's better emotional wellness. Individuals driving to workplaces announced that they were less inclined to exercise and healthy eating, telecommuters don't have those hindrances.

IV. CHALLENGES OF REMOTE WORK.

Among some of the contrary difficulties looked by the working from home workforce is the affinity to work more hours during a day. A little more than 29 percent state this is their top test. About a similar level of respondents state individuals accepting they aren't really working while they work remotely could be a significant test, and more than 26 percent refer to inspiration as their greatest test. Other big challenges for remote workers include disconnecting from work, loneliness, and bad internet connections. On the other side, some employees stated one issue with remote work was having the ability to trace employees time worked. While some people struggle with eliminating distractions and specializing in completing their work tasks, others have a tough time unplugging at the end of the work day. When you work from an office, the end of your day is when everyone goes home. Once you work remotely, it can be difficult to disconnect and step away from work when you know there are pending tasks, and although occasionally it's going to be important to relinquish the additional mile at work, making a habit of it can be detrimental to your health and cause you to burn out. If you're someone who likes indulging in gossip around the water cooler, occurring lunch dates, and meeting new people around the office, working remotely might not be for you. While you'll still meet with and interact with clients, many of those conversations take place over phone calls or video chats—not a conference room or coffee shop. To be sure, those living the cubicle life often complain of distractions within the form of regular interruptions, but sometimes distractions are preferable to isolation! If you're working remotely for yourself, you'll be tasked with finding leads, pitching, and shutting prospects, then doing the particular work, taking care of finances, and managing your time—all without the oversight of a boss or the help of another department.

V. TIPS FOR SUCCESSFUL REMOTE DEVELOPMENT

1- Hire trustworthy people to work remotely

In case you're addressing whether you can confide in your group to work remotely, you should be addressing whether you made the correct decision by employing them. The primary question to pose when hiring people is 'do I trust you to work remotely' and if the response is 'no' the meeting stops there. Regardless of whether your group is office based or not, just recruit individuals you trust to work remotely, independently and still reach the goal.

2- Make uncompromising collaboration and communication Every single effective group needs consistent correspondence and joint effort and remote teams are no exemption. You have to double-check that everybody in your group comprehends why they have to impart and work together with their associates and watch out for any individual who's especially calm. Make discussion as open as possible, stick to open channels of communication to give that buzz of work and progress you'd get if your group were in an open office.

3- Provide the means to make it happen

In the event that you need a remote group continually imparting and teaming up, you need to give them the tools to get it going. At Hired By Startups, we utilize an assortment of interior and outside apparatuses however the primary ones are Slack (for all inner correspondence),

Google Meet (Hangouts – for video calls), Google Docs and Sheets and Trello.

4- Daily Updates

Perhaps the greatest danger of building a remote group is that everybody is engaged in their own work with zero ability to see into what others are really going after. To battle this, we have a day by day standup (on Google Meet) where everybody examines their earlier day's accomplishments and focuses for the day. It's quick and reminds everybody that they are a team. After the standup, the group is urged to catch up

straightforwardly with one another on explicit questions to guarantee there's a steady progression of data and everybody is completely updated on what they are taking a shot at as an organization.

developer can remotely work from. It employs a technique that has a business model, a product road map, and a view of partners, competitors, and customers. The product is the result of the strategy. Products constantly change (improve). Strategy changes occasionally (evaluate). Vision rarely changes. Normally management, failure to deliver results is caused by failure to plan or failure to execute. Both are frowned upon. But within the modern economy, both are useful tools for testing new ideas. It carries with it 3 major stages namely: Build, Launch, Improve. We will have a look at each of these steps thoroughly.

5- Create ways to meet in person

Getting time together and getting to know other employees is indispensable. You can have the best remote group on the planet yet nothing beats some time in person. You can organize a few days a month where you come together and collaborate in person. For groups that are more geologically disseminated , you should attempt to meet for at least one week of the year, regardless of whether that is in one go or maybe two collective ends of the week.

1.Build:

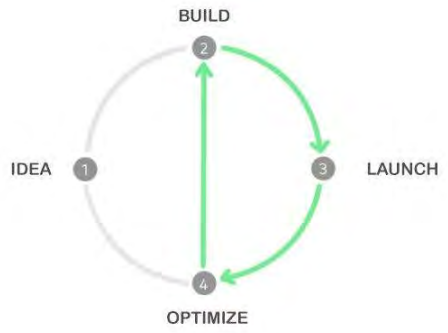


Figure 2: Development cycle of R2D

6- Get to know each other personally

Your remote group will be more productive when they truly know each other personally – passion, leisure activities and non-professional life. As far as we can tell, the best two mechanisms for personal conversation are food and drink. There can be days where the entire team sit together for some essential "team-building". If gathering week after week is beyond the realm of imagination, you could take a stab at doing this over video chat (Google Meet) – it'll likely work similarly!

7- Smart work strategy and management

The most important thing gained from driving a remote group, it's that the indispensable elements for progress (trust, correspondence, coordinated effort) don't simply occur – they're the result of a well-executed remote working technique and management. You need to show others how its done and put forth a cognizant attempt to support remote working if it will work for your startup.

VI. WORKING:

R2D method proposes incremental and iterative approach to software design which is broken into individual models that

- Understand Requirements:

Realizing an approach to assemble prerequisites might be an aptitude that each investigator, and venture director, – ought to have. Nonetheless, it is by all accounts an expertise that is commonly ailing in numerous associations. Poor prerequisites social occasion might be a significant clarification for venture issues in numerous associations. Social event necessities are over simply posing a few inquiries and afterward continuing to the accompanying advance inside the lifecycle. The essential piece of the fast remote advancement requires welcoming the correct individuals to share business objectives, innovation capacity, and client need. The objective of this stage is to extend the comprehension of the item or undertaking. This is the place the prerequisites are first accumulated. To inspire precise prerequisites, the investigator must pose the correct sort of inquiries and afterward listen cautiously to the appropriate responses. There are various strategies for inspiring prerequisites and your task may need to utilize different methods relying upon the conditions. This incorporates interviews, encouraged sessions, models, polls and that's only the tip of the iceberg.

- Distribution:

It is counterintuitive, but smaller batches are far better for progressing through a project quickly and efficiently while working remotely. They seem inefficient but allow faster turnaround for the product resulting in a more rapid iterative cycle. It helps in earlier detection of a controversy additionally as quick feedback from the customers. Toyota utilized the little bunch way to deal with contend with its somewhat more promoted American partners whose cluster sizes were generally greater.

- Evaluation:

This framework is focused on rapid software development while working with remote employees, so the progress evaluation of teams is done on outcome basis. THere is no track of how many hours an employee has worked. Only the outcomes achieved are used to measure the progress of remote teams, it develops an outcome based state of

mind which is useful for developers for long run by keeping the milestones in mind which is self-motivating. Weekly code review and progress evaluation is conducted by using various tools like Chat and collaboration tool (Google Hangouts, Slack, etc.), Video calling, Screen sharing, Project management system (Asana, Wrike, etc.), Automated onboarding software, HR software with employee self-service, Recognition and reward system. Putting processes in place and managing projects in a way that you can involve remote employees easily is an easy way to manage teams remotely.It's extremely hard to feel like you are passing up a great opportunity since you're not in the workplace. What causes the separation between colleagues to feel like separation is the absence of correspondence. On the off chance that you message someone who lives hours away and they don't react for a few hours, I'm going to wish we worked in a similar room so we didn't need to manage that. Be that as it may, on the off chance that each time I message them, they react quickly, at that point it's simply equivalent to in the event that we were both in the workplace together.

2.La
unch
:

Every business plan begins with a collection of assumptions. It spreads out a procedure that acknowledges those assumptions as ensured and keeps on pointing out a way to deal with achieving the association's vision. Since the doubts haven't been wind up being substantial (they are assumptions, taking everything into account) and in truth are often erroneous, the target of a startup's underlying undertakings should be to check them as quick as could be normal the situation being what it is.

- Minimal prototype:

A minimal prototype helps developers start the method of improving as quickly as possible.It isn't really the littlest item possible, however; it is basically the quickest method to get past the remote fast improvement cycle with the base measure of exertion. In spite of conventional advancement, which ordinarily includes an extended , attentive timeframe and takes a stab at item flawlessness, the objective of the negligible model is to begin the strategy for improving, not end it. A negligible model is implied not simply to answer item plan or specialized inquiries. Its will probably test central business theories.

- Prototype complexity:

Least model range in unpredictability from incredibly straightforward smoke tests (minimal in excess of a commercial) to real early models total with issues and missing highlights. Deciding exactly how complex a minimal prototype needs to be cannot be done formulaically. It can be done by judging different factors like urgency of deployment, lack of technical requirements from customers or simply faster optimization.

3.Opti
mize:

Before development, a project is little more than a model on a piece of paper. The financials in the marketable strategy incorporate projections of what number of clients the organization hopes to pull in, the amount it will spend, and how much income and prot that will prompt. It’s an ideal that’s usually far from where the product is in its early days. In optimization, we thoroughly measure where it is at this moment, standing up to the hard facts that appraisal uncovers, and afterward devise trials to figure out how to draw the genuine numbers nearer to the perfect reflected in the undertaking plan.

- Baseline metric:

Once the minimal product has been established, the developing teams can work toward the second learning milestone: optimization. Each advancement, advertising, or other activity that a startup attempts ought to be focused at improving one in everything about drivers of its development model. For instance, an undertaking group or organization may invest energy improving the vibe of its item to make it simpler for new clients to utilize. This assumes the enactment pace of late clients could be a driver of development and that its pattern is less lower than the organization would truly like. To exhibit approved learning, the plan changes must improve the actuation pace of new clients. In the event that they don't, the new structure ought to be passed judgment on a disappointment. This can be a

crucial principle: a great structure is one that improves client conduct. It sets out with an unmistakable benchmark metric, a speculation about what will improve that measurement, and a lot of analyses intended to test that theory. Over time, a team that’s learning its way toward a more robust working product will see the numbers in its model rise from the horrible baseline established by the minimal prototype and converge to something just like the ideal one established within the software project design plan.A startup that neglects to do so will see that perfect retreat ever more remote into the separation. At the point when this should be possible right, even the preeminent ground-breaking reality mutilation field

won't have the option to conceal this basic certainty: in case we're not moving the drivers of our plan of action, we're not gaining ground. That turns into a definite sign that it's time to pivot.

- Optimization siliency:

Engineers, designers, and marketers are all skilled at optimization. As an example, direct marketers are experienced at split testing value propositions by sending a distinct offer to two similar groups of consumers so that they can measure differences in the response rates of the two groups. Engineers, of course, are skilled at improving a product's performance, even as designers are talented at making products easier to use. Notwithstanding, these apparatuses for item improvement don't work the indistinguishable route for new companies. In the event that you are building an inappropriate thing, improving the item or its advertising won't yield noteworthy outcomes. An undertaking group needs to quantify progress against a high bar: proof that a manageable business can be worked around its items or administrations. That is a standard that can be surveyed just if a startup has clarified, substantial forecasts early. Without those forecasts, item and system choices are undeniably progressively troublesome and tedious.

- Pivot or Preserve:

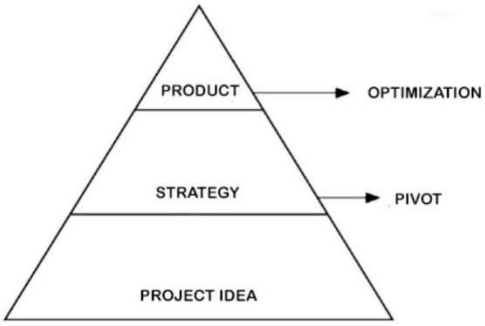


Figure 3: Pivot & Optimize

Each project team in the long run faces a superseding challenge in building up a fruitful item: choosing when to rotate and when to continue on. Everything that has been talked about so far is a preface to an apparently basic inquiry: would we say we are gaining sufficient ground to accept that our unique key speculation is right, or do we have to roll out a significant improvement? That change is known as a turn: an organized course rectification intended to test another basic speculation about the item, procedure, and motor of development. The choice to turn requires a reasonable look at and target mentality. We've examined the indications of the need to turn: the diminishing adequacy of item explores and the general inclination that item improvement ought to be increasingly profitable. At whatever point you see those side effects, think about a rotate. There are many ways to pivot based on observation and customer feedback. A popular feature can become a new product, or a less popular product may become only one of the features of the new product. Adding new features based on customer needs, Or changing product from a single-user product to a multi-user platform. New customers come from the actions of past customers. They inform others, end up showing the product to others, or end up purchasing the product again.

VII. CONCLUSION

Remote development environments, such as those provided by code-server and therefore the Google Cloud and Microsoft Azure Cloud Shell code editors, provide users with multiple benefits over local development environments. They can be connected to from anywhere, standardized across your team, and scaled to satisfy your performance needs. All of this can be accomplished with little to no changes to your existing development flow. To summarize, working remotely as a developer can be an incredible journey. Years full of professional learning, travel, connections, and new experiences. This reflection is solely based on the developer's own experience, so it can differ from someone else working remotely. People like the lifestyle and freedom they have, but it could change in the future. Their eventual goal is to become financially independent so they can spend their mental energy on things that they deeply care about. It has become the North Star that guides them forward and gives motivation. Developers see the path that can take to achieve that goal.

References

[1] Introduction to the Minitrack on Agile and Lean: Organizations, Products and Development, IEEE, 2018.
[2] Monica Rossi, Marco Taisch, Sergio Terzi "Lean product development: A five-steps methodology for continuous improvement", IEEE, 20 june, 2012.
[3] Cockburn, A. (2006). Agile software development: the cooperative game. Pearson Education.

Home Automation

Raj Jaiswa, Rajesh Gupta, Shubham Gupta
raj.jaiswal053@gmail.com, rajeshgupta828230@gmail.com, shubhamssg19@gmail.com

Abstract— **IoT, i.e. The main objective of this project is to develop a home automation system using an Arduino board with Bluetooth being remotely controlled by any Android OS smart phone. As technology is advancing so houses are also getting smarter. Modern houses are gradually shifting from conventional switches to centralized control system, involving remote controlled switches. Presently, conventional wall switches located in different parts of the house makes it difficult for the user to go near them to operate. Even more it becomes more difficult for the elderly or physically handicapped people to do so. Remote controlled home automation system provides a most modern solution with smart phones. In order to achieve this, a Bluetooth module is Interfaced to the Arduino board at the receiver end while on the transmitter end, a GUI application on the cell phone sends ON/OFF commands to the receiver where loads are connected. By touching the specified location on the GUI, the loads can be turned ON/OFF remotely through this technology. The loads are operated by Arduino board through opto- isolators and thyristors using triacs.**

Keywords – Home Automation; Internet of things(IoT); Smart home; Arduino; Microcontroller

I. INTRODUCTION

or automatic control of a few lights, or it can be a complete system that controls all major parts of your home, custom set to your own personal preference. Home automation is anything that gives you remote or automatic control of things around the home.

The concept of home automation has been around for a long time, especially in sciencefiction that imagined future homes being completely automated and robots helping us with household chores. The technology hasn't been available to make these fantasies into reality - until now! Before you can have smart appliances, you have to invent the appliances themselves! Between 1901 and 1920 there was a rapid development in domestic apparatus, kicked off by the invention of the engine-powered vacuum cleaner. After this, came the advent of refrigerators, washing machines, irons and toasters to revolutionise the lives of domestic maids - but these were still too expensive for most households to afford. As technology becomes more and more affordable, and with the majority of western populations having home internet and a smart phone, smart technologies are slowly integrating into our homes. Today, the focus is on convenience, security and energy efficiency through connectivity and interactivity.

II. LITERATURESURVEY

This project is one of the important Arduinos Projects. Arduino based home automation using Bluetooth project helps the user to control any electronic device using Device Control app on their Android Smartphone. The android app sends commands to the controller - Arduino, through wireless communication,namely, Bluetooth. The Arduino is connected to the main bluetooth and relay which has five 8 channel as shown in the block diagram. These relays can be connected to different electronic devices. Home automation is anything that enables you to use your home's lighting, heating and appliances more conveniently and efficiently. It can be as simple as remote or automatic control of a few lights, or it can be a complete system that controls all major parts of your home, custom set to your own personal preference. Home automation is anything that enables you to use your home’s lighting, heating and appliances more conveniently and efficiently. It can be as simple as remote A literature review collects scholarly publications, books, dissertations, conference proceedings, and other tools related to a particular issue, study area, or theory, and provides context for a dissertation through the analysis of past research. Research tells a story, and existing literature helps us figure out where we are in the story at the moment.

A literature review surveys scholarly articles, books, dissertations, conference proceedings and other resources which are relevant to a particular issue, area of research, or theory and provides context for a dissertation by identifying past research. Research tells a story and the existing literature helps us identify where we are in the story currently. It is up to those writing a dissertation to continue that story with new research and new perspectives but they must first be familiar with the story before they can move forward.

Analysis: Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation. Planning: After analysis we will first study about it and do some research on it for our better understanding of the project and also getting a rough picture about what would be our problem definition for the particular project. Designing: Then we will construct the design of the project and according to that, will list down all the requirements needed for the construction for the prototype of our project. Implementation:

After acquiring the requirements we first developed the connections of the hardware components (Arduino, MQ2 sensor, DHT11 sensoretc) and then added the coding part later with help of Arduino data cable to the Arduino.

Phase 2: Testing andDeployment.

Testing: After the prototype is ready we will first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding to it and will check again. Deployment: After, complete integration and testing of

project real time running and operation of the system will be done. User need to check the value.

Arduino uno

The Arduino Uno is an open-source microcontroller board based on the Microchip ATmega328P microcontroller developed by Arduino.cc. The board is fitted with digital and analog input / output (I / O) pin sets that can be interfaced with different boards of expansion (shields) and other circuits.The board has 14 Digital pins, 6 Analog pins, and can be programmed with the Arduino IDE (Integrated Development Environment) via a Type B USB cable. It can be powered by a USB cable or an external 9-volt battery, although it supports voltages from 7 to 20 volts.

III. Hardware Tools PROJECT PLANNING (RESOURCES, TOOLSUSED, ETC)

Quantity	Hardware Components
1	Arduino uno
1	8 channel relay module, 5v power supply
1	hc-05 bluetooth module, Any smartphone
12	Jumper wires

Software Tools:

Fig 3.1:Arduino Uno

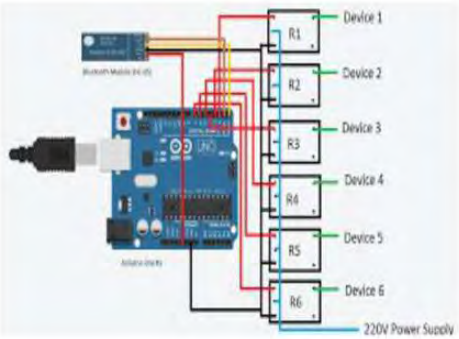
Relay circuit(8 channel):

Relays are switches that open and close circuits electromechanically orelectronically. Relays control one electrical circuit by opening and closing contacts in another circuit.. When a relay contact is Normally Closed (NC), there is a closed contact when the relay is not energized.

Bluetooth module (hc 05):

(Serial Protocol) module, designed for transparent wireless serial connection setup. HC-05 Bluetooth module provides switching mode between master and slave mode which means it able to use neither receiving nor transmitting data.

Circuit Diagram:



IV. DESIGN AND IMPLEMENTATION

- Connecte jumper wires and assemble adruino using circuit diagram.
- Connect Bluetooth module.
- Connect smart phone with bluetooth module and install application.
- Connect electrical appliance to the Relay module.
- Once done above steps, you can operate various appliances.

V. RESULTS

The adruino is given a code to operate on commands recieved via bluetooth module. This bluetooth module is connected with a smart phone. The smart phone has

installed an application to control 8 equipments. Using the application user can send commands which will be received by HC-05 bluetooth module which will execute instruction on adruino and gives us desired results.

A low cost and efficient smart home system is presented in our design. This system has two main modules: the hardware interface module and the software communication module. At the heart of this system is the Arduino Mega 2560 microcontroller which is also capable of functioning as a micro web server and the interface for all the hardware modules. All communication and controls in this system pass through the microcontroller

VII. IMPROVEMENT SINRE SEARCH G APS (FINDINGS)

- Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. They are as follows: Proposals are made to design and build intelligent home meeting the modern requirements satisfying most living standards of consumers. Home security activities supported by real time audio visual system which

  permits only the entry of authorised visitors into the flat are incorporated in the system. Easy handling of selected home appliances in the flat by switching ON or OFF from any room is provided in the system. Moreover, remote control of these selected appliances are performed through telephone dialling or with Internet are accommodated in the system. A Home Assistant system software installed in the PC brings the daily activities to the monitor screen as obtained from the activities programmed in tables of data bases.

- The electrical energy requirements of the flats are partially supported by a unit extracting solar energy and wind energy. Saving of electrical energy by this unit is maximised by automatically changing the angles of the solar panel and wind mill fixed on the terrace. Furthermore, intruder detection unit included in the system dials automatically a pre-programmed sequence of digits to give message to a remote location while triggering a local alarm.

VIII. FUTURESCOPE

Home of the future is a space for the digital natives. With the invention of lots of automation technologies featuring IOT and AI, home automation has become a reality. One can implement several of their tasks with just a single command of verbal instructions. These technologies can used to build fully functional home automation system and control smart home devices including smart lights, connected thermostats, and appliances.

• Smart spaces outside homes: Smart parking through sensors will help to recognize whether the parking is available or not. Camera monitoring can be done and with the help of artificial intelligence and computer vision, both parking facilities and security can be provided. It would be a faster and smoother process and act as a reference for other smart systems to be build accordingly. Streetlights can also be automated through sensors and build for effective use for the people nearby.

Personal home delivery: Drones will be used to deliver the packages at the right time. They will replace the normal salesman job. They might also be used for several other tasks like monitoring the weather outside the home, returning something back to a relative's home nearby and so on. They can also be used for monitoring the traffic in our locality

The home automation system has been experimentally proven to work satisfactorily by connecting sample appliances to it and the appliances were successfully controlled from a wireless mobile device. We learned many skills such as soldering wiring the circuit and other tools that we use for this project and was able to work together as a team during this project. The Bluetooth client was successfully tested on a multitude of different mobile phones from different manufacturers, thus proving its portability and wide compatibility. Thus a low-cost home automation system was successfully designed, implemented and tested.

References

The web sites that provide the information:
[1]. www.autogates.com.my
[2]. www.usautomatic.com
[3]. www.amazinggates.com
[4]. www.microchip.co.uk
[5]. www.quora.com
[6]. www.engineeringprojects.com
[7]. www.efy.com
[8]. www.electronicsworldforyou.com

Automatic Medicine Reminder Machine

Akash Jangid, Asmit Juneja, Bhaarat Kotak

Thakur College of Engineering & Technology

jangidakash286@gmail.com, asmitjuneja333@gmail.com, kotakbhaarat@gmail.com

Abstract— **IOT, i.e. Internet of Things, is the currently the most trending field in the Branch of Information Communication Technology Recent years have seen a rising in wearable sensors and today several devices are commercially available for personal health care and activity awareness. A recent health care system should give better health care services to people at any time anywhere in an affordable and patient friendly way. Currently, the health care system going to change from a traditional approach to a modernized patients centred approach. In the traditional way the doctors play the major role. For necessary diagnosis and advising they need to visit the doctor. There are two basic problems related to this approach. Firstly, the health care professionals must be in place of the patient all the time, the patient remains admitted in the hospital, wired to bedside biomedical instruments, for a long period. In order to solve these two problems the patient oriented approach has been received. Recent information from United Nations predicted that there will be 2 billion older people by 2050. In addition, research indicates that above 89% of the aged people are likely to live independently. However, medical research found that above 80% of the aged people older than 65% suffers from at least one chronic disease making them to have difficulty in taking care of themselves. Accordingly, providing a decent quality of life for aged people as become a serious social issue at that moment. The rapid proliferation of information and communication technologies is enabling innovative healthcare solutions and tools that promise in addressing the above challenges. Now, Internet Of Things (IOT) has become one of the most powerful communication paradigms of the 21st century. In the IOT environment, all objects in our daily life become part of the internet due to their communication and computing capabilities. Heart rate is one of the fundamental physiological limits, essential for monitoring and diagnosis of patients. To keep people effective and healthy, a readily accessible modern health care system is proving to be effective in saving costs, reducing illness and prolonging life. In this paper, an enhanced healthcare monitoring system is described, that is smart phone based and designed to offer wireless approach and social support to participants. After studying the literature survey it was planned to overcome the research gaps and the system was tested using the test cases and then the outputs were generated. The output of the gas detectors were been shown according to different test cases and thus the gas detection and sms alert system become more responsive then those currently in use**.

Keywords – Internet of things.

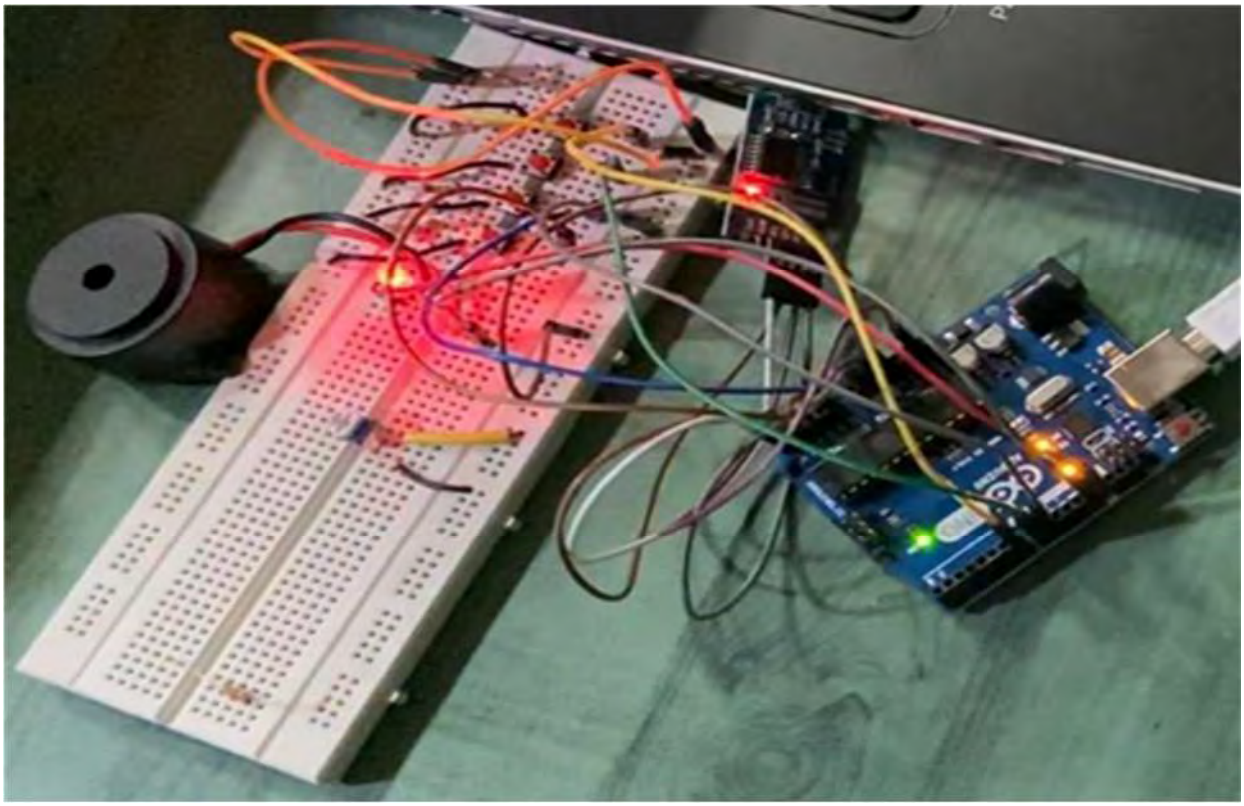
I. INTRODUCTION

Many medical errors are due to the fact that people in charge of patient or elders medication have to deal with sorting huge amounts of medicine each day. There are a lot of people who actually need help on daily basis – may it be our elderly people, family members, relatives, the ones who have special needs. Elders are more affected by the timing of taking a certain medicine as compared to others, in order to prevent any dysfunction or illness timing is a must. Automatic Pill Reminder Bottles Invented and patented in 2004 by innovator and inventor Joseph Lai the pill reminder bottle. This device is compatible and can be retrofitted inside a regular conventional pill bottle cap hence making it easily portable. This reminder device is installed inside the conventional pill bottle between the bottle cap and the bottle container. When the user closes the pill bottle cap on the bottle container, the electronic timer, with factory predetermined time interval, is automatically activated. That activated timer will generate alert signals not only remind user last pill has taken but also to remind the user to take his/hers next medicine on timely baisis. David P. Wagner invented the first pill organizer.Wagner was prompted to invent the device when he and his wife had trouble remembering if she had taken her daily pill. The application of Smart Medicine Reminder machine can be used worldwide at any time irrespective of the climatic conditions. When it comes to our loved ones, we always want to them to stay fit and healthy. But what will happen if they get ill and forget to take medicine on time. We become worried when people who are close to us forget to take medicine on time. At hospitals, there are many patients and it is difficult to remind every patient to take medicine on time. The traditional ways require human efforts to remind to take medicines on time. The digital generation doesn't follow that and we can use smart devices/machines to do that. The application of Automatic Medicine Reminder is a smart, very wide and can be used by patients at home, doctors at hospitals and at many other places. When it comes to reminding it is accurate and fast. Being light in weight it is easily portable.

II. SUBJECT DESIGN AND CURRICULUM

Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation. After analysis we will first study about it and do some research on it for our better understanding of the project and also getting a rough picture about what would be our problem definition for the particular project. Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the

solution is technically feasible. As part of this, the costs and benefits associated with the proposed system compared and theproject is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

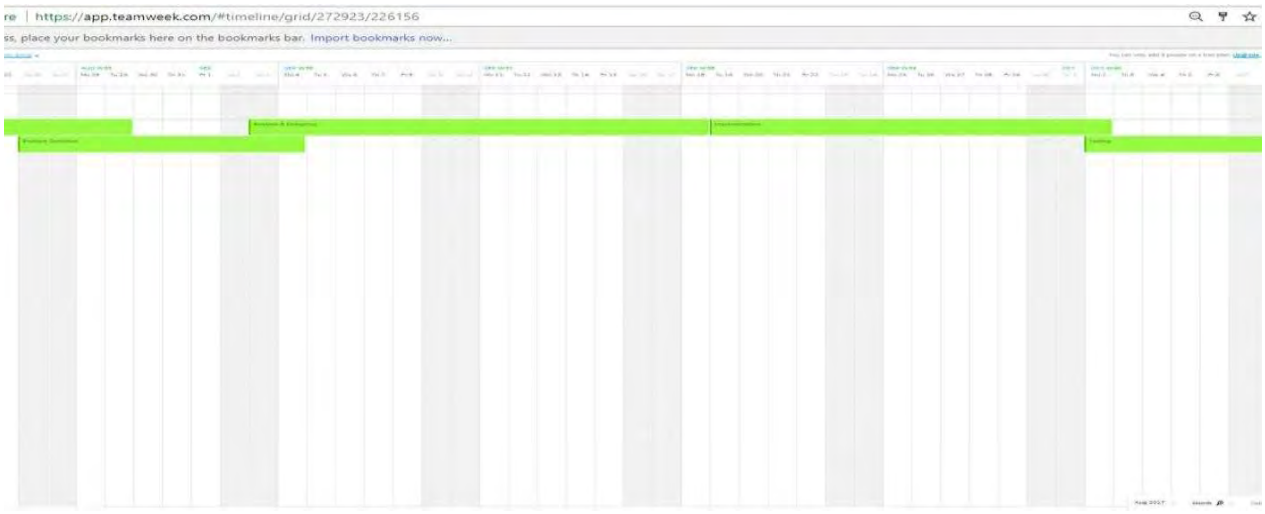


A. Subject Learning Outcomes and Contents

There are many systems which are helping for the same purpose. But these systems are problematic to use, non-mobile, expensive and complex process. The proposed system overcomes these problems. The 8051 Based Medicine Reminder is simple to use, affordable, better accuracy. This system is co-operative for every age group and can also be used in hospital for a group of people. This system will definitely reduce the bad effect caused due to wrong intake of medicine. This project in future may be more desirable via integrating it with GSM technology, in order that the affected person gets a reminder approximately the medication he has to take via SMS on his/her mobile phone. Also a provision to change the name of the drugs can be included by means of interfacing the device with a PC or EEPROM (non-volatile memory).

B.Subject Delivery

Table 1-Gantt Chart



The proposed method takes an automatic control action upon detection of time intervals of medicine. Initially if a delay is detected then the electronic sensor i.e. the buzzer that obeys the principle of sensor senses the delay, if any delay in timing is sensed then the output of this sensor goes high. This high signal is monitored by the microcontroller and it will identify the amount of delay occurred in taking the medicine . If there is a delay , the consumer is informed about it through the LED light and Sound of the buzzer.A Gantt chart is shown in Figure 1.

C. Components required

The Arduino UNO is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc.[2][3] The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits.[1] The board has 14 Digital pins, 6 Analog pins, and programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable.

The **DS3231** is a low-cost, extremely accurate I2C real-time clock (**RTC**) with an integrated temperature-compensated crystal oscillator (TCXO) and crystal. The device incorporates a battery input, and maintains accurate time keeping when main power to the device is interrupted.

GSM/GPRS module is used to establish communication between a computer and a GSM-GPRS system. It requires aSIM (Subscriber Identity Module) card just like mobile phones to activate communication with the network. Also they have IMEI (International Mobile Equipment Identity) number similar to mobile phones for their identification.

LCD modules are vey commonly used in most embedded projects, the reason being its cheap price, availability and programmer friendly. Most of us would have come across these displays in our day to day life, either at PCO's or calculators. The appearance and the pinouts have already been visualized above now let us get a bit technical. **16×2 LCD** is named so because; it has 16 Columns and 2 Rows. There are a lot of combinations available like, 8×1, 8×2, 10×2, 16×1, etc. but the most used one is the 16×2 LCD. So, it will have (16×2=32) 32 characters in total and each character will be made of 5×8 Pixel Dots.

III. SCOPE OF THE PROJECT , UNDERSTANDING ,RESULT AND DISCUSSION

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too.As this system is currently not Wi-Fi based also there is no message to doctor via sms or email. This system only rings when there is time to take the medicine ,basically it rings after particular intervals of time but cannot show the exact time as shown in a clock. Also, the Potentiometer The major disadvantage is that it requires a large force to move their sliding contactsFurher we can use Lcd screen and also In Built voice Command which will make the system more accurate and attractive.

There are many systems which are helping for the same purpose. But these systems are problematic to use, non-mobile, expensive and complex process. The proposed system overcomes these problems. The 8051 Based Medicine Reminder is simple to use, affordable, better accuracy. This system is co-operative for every age group and can also be used in hospital for a group of people. This system will definitely reduce the bad effect caused due to wrong intake of medicine. This project in

future may be more desirable by making use of LCD and also by inserting In Built Voice Commands . These changes will make the machine more advanced as well as attractive.

A. UNDERSTANDING

In this after understanding the topic the project feasibility was analyzed by performing different types of feasibility studies and by also planning the project tools, their project schedule, timeline charts, etc. Feasibility study will help in better understanding the various feasibilities associated with the project and helping to make the correct decisions and completing the project within the schedule, budget, etc.

The tools were specifically identified in this chapter stating which technology can be feasible and how conveniently the project can be completed. This helps to understand the technology and tools that can be used for the project. The Gantt chart helps us to track the project and see the schedule of the project and to see if the project is on the right track and on schedule and not behind the deadline

B. RESULT AND DISCUSSION

Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase.

Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done.

Proposing design and architecture for the prototype.

Implementation of the Arduino model with the help code running it in the Arduino IDE environment.

Testing was done with cases like medicine taken to test the buzzer, taking the medicine on time and the testing was done in a clinic under doctors guidance.

Real time sensing was done when in the doctor's clinic when a patient had to take the medicine and it was already time for him to take it and the buzzer reacted

IV. CONCLUSION

We have successfully shown the Medicine Reminder Machine by IOT based project is possible that patients can take medicine on time and the alarm so loud such that even the sound of it can be heard in areas where huge amount of noise is produced. It helps doctors, nurses and the caretakers for speedy recovery of the people .Being cost effective and light in weight it can be useful to everyone in the society.

ACKNOWLEDGMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have completed the project successfully. I would like to thank everyone for their guidance.

We sincerely thank our principal, Dr.B.K.Mishra and HOD, Dr.Rajesh Bansode for always encouraging us to do our best. We highly indebted to our guide Dr. Sangeeta Vhatkar, who supported and constantly supervised us throughout the project and helped us in not only completing this project but also provided us with ample amount of knowledge that was really beneficial to us.

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of IT Department who helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in laboratory for their timely support.

We would like to express our gratitude towards our parents for their kind co-operation and encouragement which helped us in completion of this project.

REFERENCES

1. Samir V Zanjal,Girish R Talmale Medicine Reminder and Monitoring System for Secure Health Using IOT,

Vol. 1 Issue 9, September 2016.

2. Jabeena, Animesh Kumar Sahu, Rohit Roy, N.Sardar Basha Automatic Pill Reminder For Easy Supervision, Vol. 2 Issue 2, July 2017.

3. Nitesh P. Sonawane, Vijay D. Chaudhari, Dr. K. P. Rane. Automatic medicine reminder with RTC interface through mobile and whatsapp, Vol. 1 Issue 3, June 2016

4. Milan Ramljak Smart home medication reminder system, Vol. 1 Issue 2, July 2017

5. R. Surender Reddy , C. Niharika , Rohit Sharma , A. Bharath Kumar ,E.Narayana. An Intelligent Patient Medicine ReminderSystem,Vol.3,Issue8,August2017.

Student Grievance Support System

Bimalesh Seth, Darsh Shetty, Vidhi Rana, Vikas Tiwari, Hardik Sodhani, Kavin Parikh
Department of Information Technology, Thakur College of Engineering and Technology, Mumbai
sethbimalesh@gmail.com, hettydarsh002@gmail.com, vidhidrana20@gmail.com, v8kgsm@gmail.com, hardik.sodhani@gmail.com, kpk1611@gmail.com

Abstract – **The goal is to design an online system which fetches the grievances from the users, categorizes it in a proper headings/tags and shows it to the admin. The complaints are safe/secure and the identity of the complaint is not revealed to any other user. The Online Grievance management system also shows the status of the complaint and the kind of action taken on the same. Administrator can generate a report of this system in between selected dates of his own choice. The data given by the users is secured and safe as the application and the database is protected by a 64 bit encryption, preventing the identity of the user to be revealed. This helps the user to prevent getting into further trouble. Only the authorized access to the reports will be available which just shows the time and date of entry by the user**

Keywords - Validity, student's grievance, process, student/ user.

I. INTRODUCTION

A grievance is an emotion that a person has about a current existing system. Most people don't come open with the problems they have and just keep it in their mind. Thus our objective was to create a way so as to make the people to come open about the problems they have with the system and we came up with a solution of an online Student Grievance management system which can be used to submit complaints regarding their problem online enabling the user to upload or post his complaint from anywhere by using an application on his mobile phone, computer or any other device that has access to an internet connection. This application will act as a safe platform for the people to tell their grievances without the fear of getting judged or being pushed into deeper trouble. This system also saves a lot of time when compared to the traditional manual system, saving the efforts and time of both the sides i.e. the administrator and the users.

II. HARDWARE REQUIREMENTS

A system which is compatible to support internet

III. SOFTWARE REQUIREMENT

Browser - Mozilla, Chrome, Express.js, MongoDB

IV. SCOPE

SGSS will be helpful to receive problems which are reported by the students for any institute. This platform will be extremely helpful for the Institute's management to understand what kind of problems are faced by the students. This system will help us to give quick responses to grievance and help to save time which is a very crucial requirement to any institute.

V. LITERATURE SURVEY

Initially, we have studied the current system available for grievance collection in college and also met administrative staff and the principal of college for collecting their requirements. After collecting different requirements of staff, we have met various students of different branches and took their suggestion and asked them what kind of system they want for hassle free resolution of their queries. We have also gathered data about challenges that are faced by students before submitting their grievances and after submitting their grievances. There can be two types of systems in which grievances are submitted either with a person's identity or anonymously. For submission of Anonymous grievances, the system must have high security so that none but the admin is able to see the student's identity. Security of this system ensures the student for their privacy and fearless submission of grievances. College authority can easily and quickly resolve student grievance through this online platform. Our team found out that the most appropriate to the research topic handling customer grievance using Android Architecture was tried to improve relation between students and faculties by presenting a new model based on Service Oriented Architecture (SOA). it will involve user in decision making and at the same time it will help the authority in reducing the people problems and dissatisfaction. E-Complaint, regarding this each organization has its own definition for grievance. Some define grievance as something related to the services they provide for users. The value of complaints, both as a bridge that connects the customer and administrator and as a means of giving them a chance to turn a dissatisfied customer into a satisfied and loyal customer. Customer complaining behaviour can be defined as a result of customer dissatisfaction. On the other hand, customer satisfaction is not an absolute scenario, but very much depends on interactions, feedback, understanding, and reviews. Grievance Redressal System is a system in which we accept the user grievance and it will make sure that the user grievances are properly solved and managed by the particular department to which grievance belongs. Based on this is the basic idea on which the grievance and report system of multinational brands like Amazon, Flipkart, Walmart function.

According to study in [1], unhappiness leads customers to both migration behaviour and negative referrals to other potential buyers, adversely hitting retention rates, profitability and organizational goodwill. Therefore, customers who make grievances provide the company with additional chances to identify internal shortcoming and to establish a suitable recovery strategy. This should focus on

gravely encouraging complaining actions and establishing influential procedures to solve grievances, remove the causes of dissatisfaction, and provide result in answers to affected customers. Accordingly, by designing an integrated complaint management system, organizations have the opportunity to learn from customer feedback and to use this information in order to reduce weaknesses, enhance business performance, avoid future opposing experiences, and necessarily re-establish customer satisfaction, faithfulness and relationship commitment.

Students are the most vulnerable parts at educational institutions often unable to express and sometimes fail to attempt proper support for the issues they face appearing at numerous levels. A prototype of grievance redressal had been worked out which could meet well with the solution delivery for the occurring conflicts for students. In this paper, The focus is on the evolution and the implementation of the above-mentioned prototype which could be integrated to be fixed to the grievance redressal for students. That paper puts deep perception into incorporating all those problem aspects which were noticed on the basis of the analysis phase plus some additional needed areas. [2]

In [3], Online students' complaint is an information system which is able to help study program in accepting censure and suggestions by the students to help it enhance the services. This creation will reduce paper usage, time and energy. This research implemented prototype model as development system model to structure, plan and design the system. The Complaint Management System [4] gives the service to the Users to see their previous grievances and also, they can see the remark provided on every stage of their grievance. As per the security reason this system needs an email validation which means only one email can be used by one user to use. For Admin the extra resources provided are that he can smartly track the User and it also displays the attempts taken by user to login into system and it gives the Email notification to the Admin on his Email about every new grievance.

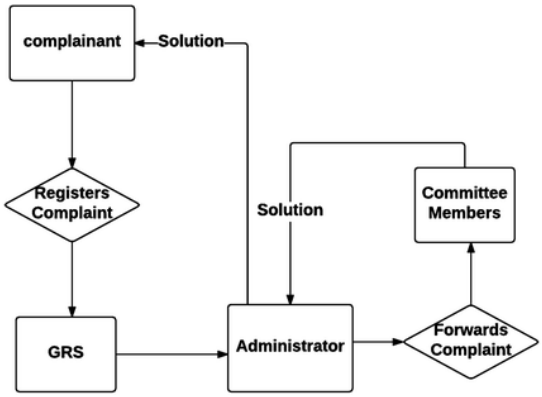


Fig. 1. System process

This website can be easily used by people even by those who are not from IT background. It is much easier and secure way of submitting a complaint.

This project includes two modules.

User to Admin: User can create their accounts and can complain about their problems and also can see action taken on their complaint. Users can check the status of current and as well as previous complaints. Users can submit their complaint easily after a quick registration. Existing users can lodge a complaint by using their registered username and password.

Admin can check the complaints also can take action on it. Admin also has the authority to remove the user. Admin can check the monthly report generated by the system. New features can be added as per requirements.

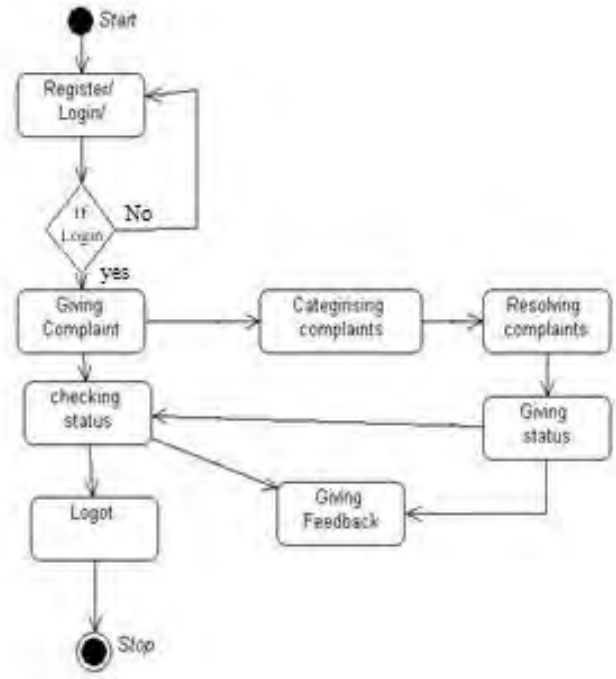


Fig. 2 Flowchart of model processing

VI. PROPOSED SYSTEM

The system which we have proposed that contains user interface through which a user can convey a grievance which is supposed to reach out to authorities hierarchically such that an action in accordance with the grievance can be taken as soon as the grievance is submitted by the user. Our system works in such a way that the user can easily access it in a user-friendly manner so that the grievance come into action. There are two types of grievance submission sections through which a user can submit the grievance with or without his/her personal details. It can be used widely within college campus, offices, industries, etc and not restricted with any particular area. For reference of system flow See Fig. 2.

VII. PROS AND CONS

The pros and cons of the proposed GRS system primarily includes human negligence since the system, though an automatic prototype to redress the complaints of the students/victims, will be controlled by humans who could be negligent at times in forwarding complaints to the respective committee member or while providing solutions to the respective administrator. Another shortcoming includes a poor network which could persist at times; as a result, the system of forwarding and resolving the complaints in the form of sending back the replies to the respective administrators and then to the concerned student could get delayed, the proposed GRS system being based on simple mail transfer mechanism.

VIII. FUTURE ENHANCEMENTS

The GRS working on the pretext of the grievance redressal for all students currently works as a Web application among the various members and the targeted audience. To extend this further in order to fulfill various requirements, following enhancements are suggested: Though several future enhancements of the system worked upon are possible, the prime focus also consists of development of a mobile application in order to increase the mobility of the application since the future limits the usage of mobile applications and as seen portable devices are found everywhere which will facilitate the receiving of all the notifications in the cell phone by the members. A Prototype for Grievance Redressal System and students associated with the application further increasing the reliability of the system and the rate of problem-solving at the same time. The mobile application and web application is targeted to improve the user experience by providing them with additional features for uploading the pictures the proof in the form of audio or video files, which might improve the case solving capability especially in such cases with a high rate of severity or even fatality. A toll-free helpline could be made available on a 24by7 basis for the victims in order to register a complaint at emergency hours or to seek counsel in case of catastrophes.

IX. RESULT AND DISCUSSION

This technical paper is an attempt to enlighten the fact that there are hardly such systems prevailing curtailing to the complaint redressed for students enrolled in numerous organizations. This article has demonstrated a GRS system for the grievance redressed of students covering various domains of complaints which could be lodged and solved easily as compared to the traditional system which involves paper and writing. The technologies used comprise of HTML and CSS to design a user-friendly graphical interface. PHP and SQL to keep track of the records at the back end. This system would be enhancer for any organization for the resolution of complaints of students and thus lead to a qualitative and quantitative development of the organization.

X. CONCLUSION

The Student Grievance Management System acts as a time saviour for the organization and helps them solve the issues faced by the students keeping the details of the student safe and secure. This helps save paper which was used in the traditional system making management of the report simpler and easy to access and solve for the admins. It was revealed in this study that organizational issues are inevitable. A grievance management system for the Institute that receives various complaints from students. Complaints lodged in the sectors like Academic, Administrative, social and other issues relating to the student. This platform allows for complaints to be lodged remotely by students with issues relating to their registration, examination, examination result, computation of their grades and hall of residence complaints and thereby improves the response time for the appropriate unit to resolve the complaints. This System is a boom for the organization as it can focus on development in all aspects.

References

[1] https://expressjs.com/en/4x/api.html
[2] https://mongoosejs.com/docs/index.html
[3] Subhash, C., Ashwani K.: Assessing grievances redressing mechanism in India. Int. J. Computer. Appl. 52(5), 12–19 (2012).
[4] Dipankar, M.: Solution towards effective complaint registration system in Indian scenario. In: IJCA Proceedings in National Conference on Advancement of Technologies—Information Systems & Computer Networks (ISCON—2012), vol. 1, pp. 1–2 (2012)
[5] J. Comput. Appl. 52(5), 12–19 (2012)
[6] IJCA Proceedings on National Conference on Advancement of Technologies—Information
[7] Enayat Alkhider, Osman Nasr, "Online Complaint Management System", IJISET - International Journal of Innovative Science, Engineering & Technology, 6, June 2015.

GIRL UP- A positive psychology game for women & child development

Mohit Gupta, Rohan Kuckian, Varun Mewada, Priyal Dumralia, Sakshi Madkholkar
TCET, MUMBAI
mohitrg2000@gmail.com, rohan.kuckian@gmail.com, varunmewada1@gmail.com, priyaldumralia22@gmail.comsakshimadkholkar1210@gmail.com

Abstract— **Due to reasons like Portability, free availability and being a great stress buster, mobile games have always been part of the reason why smartphones have never gone out of trend. Games have a great quality of affecting your conscience without you even noticing about it. A lot of popular mobile games in today's day and age like PUBG, Call of Duty etc. don't send a good message to the youth. In this paper we want to present a game GIRL UP! Which is a positive psychology game which would help in women & child development in India. The situation of women & children in India has improved significantly but it is still not very respectable Positive Psychology games can influence a person's conscience by nurturing positive emotions, promoting engagement, as well as enhancing social integration and connectedness. We hope that our game can help in successfully improving the mentality of people and make our country a better place to live for women and children.**

Keywords – Positive Psychology, games, women & child development, mentality.

I. INTRODUCTION

.The development of positive psychology is to give us a new perspective. With the further study of video games, more and more researchers have begun to shift the focus to the positive field of video games and how bringing positivity into a very highly utilized commodity like video games can help in the overall development of the society.

The game will start with the Prime Minister addressing to the ministry that there is a need to implement a few reforms so that people especially in the rural areas are educated about the well-being of women & children. We plan to have a township game where initially a small town will be assigned to the player. The player will have to initially undertake a quiz according to which he/she will get some coins which they need to invest in the town to build hospitals, schools etc. A mentality meter will be allocated to every town which would check the mentality of the town with respect to women. As the mentality meter crosses 50% the player will be allowed to expand to different towns. Also, there will be an additional feature of a quiz which can be used by the user to get quick coins. Taking these quizzes will also affect the mentality meter.

.

Our game mainly focuses on innovative and intuitive ways in which we can subtly influence a person's thinking to develop a positive attitude with respect to women & children. Traditional psychology paid more attention on the negative effect of the mobile games. Exposure to violent video games will increase aggressive behavior in both the short term and the long term.

II. PROBLEM DEFINITION

Positive psychology games (Mobile/Computer) which help develop a. Respect for right and dignity of women and child b. which help promote objective of Beti Bachao Beti Padhao (BBBP).

III. BACKGROUND

According to SALEN & ZIMMERMAN (2004). RULES OF PLAY: GAME DESIGN FUNDAMENTALS **"A game is a system in which players engage in an artificial conflict defined by rules, that results in a quantifiable outcome".** It also implies that players create an artificial world in their mind when it comes to playing games. A lot of times they end up doing something in real life which has been inspired by a previous gaming experience. Hence these games manage to affect the conscience and thinking of an individual.

The researches of positive emotions in positive psychology focus on subjective well-being. Playing video games is conducive to arouse the emotion experience. Gamers had significantly better emotional experience, higher levels of well-being, and to some extent, less depression compared to non-gamers

Most of the popular games in today's day and age do not focus on building any sort of positive psychology. Either they are brainless games which serve a purpose of entertainment only or games that instill a violent or negative psychology among young people.

There are quite a few positive psychology games but none of them serve the purpose of women & child development.

IV. IMPORTANCE

As mentioned above most of the popular mobile games nowadays spread a negative psychology among the youth. One such example of a game is the very famous online multiplayer game PUBG (Player Unknowns Battle Ground). The negative effects of such games are that Children who play more violent video games are more likely to have increased aggressive thoughts, emotions,

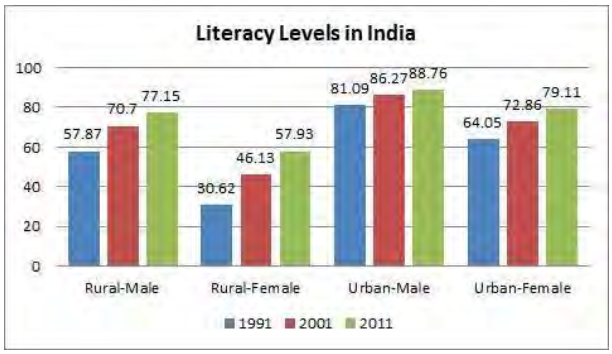
and behaviors, and decreased helping others.

On the other hand, games like Uncharted Waters collects many ports and islands in Asia, Europe, America and other places in 16th century, players in the game can find different parts of the historical sites and cultural landscapes, it widens the geographical knowledge of players. Meanwhile, to some extent, video games can improve individual's critical learning abilities and creativity. Since players need to learn the rules and the special symbolic meaning in the game, players can think critically the game as a system and a well-designed space, rather than a simple game environment.

Critical and creative thinking in the context of the game holds an important position in human psychological development and in the formation of thinking. Besides, video games have an important affect on people's behavior, moral behavior education which based on the emotion can make use of the positive effect of video games. Positive video games can promote virtue and strength of players, which is the core of positive qualities of individuals.

V. OBJECTIVES AND SCOPE

The main objective of developing this game is to improve the current situation of women & children in India There is a wide gender disparity in the literacy rate in India: effective literacy rates (age 7 and above) in 2011 were **80.9%** for men and **64.60%** for women. The low female literacy rate has a dramatically negative impact on family planning and population stabilization efforts in India.



Our game through its subtle ways wants to instill an ideology into the society that Women and Men are equally important. Due to the patriarchal approach especially in the rural areas, this game becomes that much more important as an effective way to feed a good ideology into the minds of the illiterate.

VI. PHASES

Phase1: Planning, Analysis, Designing and Implementation.

Analysis: Getting strong knowledge of the project title and doing exploration on it, we will get our description.

Preparation: After finalizing the problem statement from the SIH website we researched on existing similar games and found out the pros and cons.

Designing: Then we will build the design of the project and agreeing to that, will list down all the necessities needed for the creation for the prototype of our project.

Implementation: After gaining the requirements, we first learnt about game building software like UNITY and required back end database like SQLite.

Phase 2: Testing and Deployment.

Testing: After the model is ready, we will first associate the hardware with the allotted code and then we will check if it supports the machine or not. If not, we will resolve the issues relating to it and will crosscheck again.

Deployment: After completion of integration and testing of project, real time testing and operation of the system will be done.

VII. METHODOLOGY

The basic 2-D UI will be designed using UNITY. User details and game progress and quiz questions will be stored using SQLite database. More emphasis will be given on making the game user friendly and attractive so that it entertains as well as educates.

VIII. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware

- Any Computer system with the required software installed

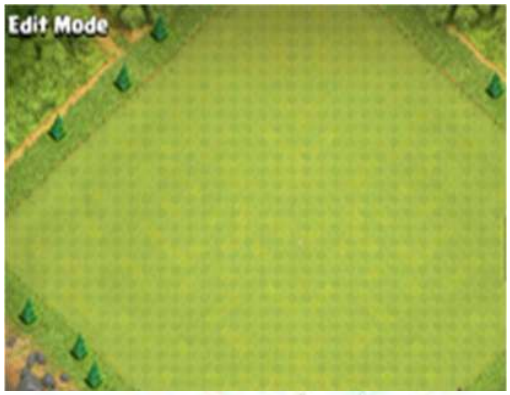
Software

- UNITY: Sandbox environment for developing the game
- C#: Used for writing the code
- SQLite: For designing of the database

For designing the game, it was planned that Unity would be used as it provides great Platform Support, very good features to add some enthralling graphics, helpful documentation for beginners to understand the dynamics that go behind making the game and also the support for easy debugging.

SQLite was chosen to design the back-end of the game as it is lightweight, reliable, portable and provides better performance as compared to other database systems when It come to dealing with comparatively smaller datasets.

IX. PROSPOSED UI SCREENSHOTS.



The above diagrams represent the proposed UI of the game. The game will start with empty land given to the player. Then the player with the help of coins has to build on that land and look to improve the mentality meter.

X. OUTPUT

The game will start with the Prime Minister addressing to the ministry that there is a need to implement a few reforms so that people especially in the rural areas are educated about the well-being of women & children. We plan to have a township game where initially a small town will be assigned to the player. The player will have to initially undertake a quiz according to which he/she will get some coins which they need to invest in the town to build hospitals, schools etc. A mentality meter will be allocated to every town which would check the mentality of the town with respect to women. As the mentality meter crosses 50% the player will be allowed to expand to different towns. Also, there will be an additional feature of a quiz which can be used by the user to get quick coins. Taking these quizzes will also affect the mentality meter.

XI. OUTCOMES

Creating the outline, works survey and achievability study for proposed solution which will act as an input to the design phase.

Suggesting Statement of work, scope description and scope border for planning the model from the problem definition to decide what needs to be done and what not to be done.

Phase 1 Implementation:

Implementation of the game with proper connectivity and no failures.

Phase 2 Testing:

Testing was done by making sure that the UI elements in the game are working smoothly and the game is not hanging or stopping due to errors in the code.

Phase 3 Deployments:

Real time working was done when we played the game until a certain level and it showed no faults. Thus, the model works efficiently.

I. RESEARCH GAPS

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. They are as follows:

- The game might take more time to load on Android phones whose version is outdated.
- The game might fail in its purpose of educating people if it becomes too predictable and the players answer quiz questions based on the amount of coins that they will receive rather than their own opinion

XII. SUMMARY

Our game GIRL-UP focuses on improving the mentality of people in our country with regard to women & children. Games have become a very important part of smartphones and they subtly manage to influence our conscience. Most of the famous games nowadays are negative psychology games which affect the youth of the country in a negative manner. Hence a positive psychology game like GIRL-UP is the need of the hour.

XIII. CONCLUSION

We have successfully shown the implementation of our game. We will be using UNITY and C# to design the front-end and UI of the game. SQLite will be used to design the back-end part because it is a lightweight and portable database system which is efficient for such a kind of a game.

References

[1]. https://www.gamesradar.com/why-mobile-games- are-so-popular/

[2]. https://link.springer.com/chapter/10.1007/978-3- 319-07464-1_16

[3]. https://content.taylorfrancis.com/books/download?d ac=C2010-0-36723- 2&isbn=9781136583094&format=googlePreviewPdf

[4]. https://www.javatpoint.com/sqlite-advantages-and-

disadvantages

[5]. http://www.potenzaglobalsolutions.com/blogs/5- rarely-known-advantages-and-disadvantages-of- unity-game-development

[6]. https://www.scirp.org/pdf/PSYCH_2012123117272 736.pdf

Poshan Abhiyaan

Rohit Das, Ritesh Dewoolkar, Ritik Gupta, Shivam Gupta, Saurabh Dubey
Thakur College of Engineering and Technology, Mumbai, India
rohitd3476@gmail.com, riteshdewoolkar@gmail.com, guptaritik349@gmail.com, smg150900@gmail.com, saurabhdubey0211@gmail.com

Abstract— **Malnutrition is not a direct cause of death but contributes to mortality and morbidity by reducing resistance to infections. There are a number of causes of death of children such as prematurity, low birth weight, pneumonia, diarrhoeal diseases, non-communicable diseases, birth asphyxia & birth trauma, injuries, congenital anomalies, acute bacterial sepsis and severe infections, etc. Our website focuses on how to improve nutritional outcomes in the country. The programme through the targets will strive to reduce the level of stunting, under-nutrition, anaemia and low birth weight babies. This website will bring clarity amongst the people, about the problems related to malnutrition and many diseases which cause death. Its benefits are amazing and this will help to reduce the negative impact of malnutrition on socio-economic development of India. We hope that our research will spread awareness amongst the people about the POSHAN ABHIYAAN. Our main aim is to promote holistic nutrition among women and children with the help of IT.**

Keywords–Malnutrition, diseases, holistic nutrition, women and child development, Poshan Abhiyaan.

I. INTRODUCTION

Our website has been specially designed to strengthen the Service Delivery System as well as the mechanism for Real-Time Monitoring (RTM) for nutritional outcomes. POSHAN ABHIYAAN was launched by Prime Minister Shri Narendra Modi in Jhunjhunu, Rajasthan in March 2018. Women from different villages and slums would be able to manage their work easily and awareness would be spread with the help of the website for efficient working. There will be different sections in which they will be provided with information such as how to feed their child, information on regular checkups and when to consult the doctor. Our website will be mainly concerned with how to stop malnutrition and also for pregnant women on what to be consumed or what not to consume during their pregnancy. The aim of this website is that till the end of the year 2020 there will be awareness spread amongst the people in every corner of India.The website will consist of honourable Prime Minister, Narendra Modi informing about the importance of Poshan Abhiyan and the need of the website to spread awareness for a disease-free and malnutrition free country till the end of the year 2022. Implementation strategy would be based on intense monitoring and Convergence Action Plan right up to the grass-root level. Our website will also concern with how to reduce stunting growth, anaemia among women and children.

II. PROBLEM DEFINITION

India is home to one of the largest populations of malnourished children in the world. Today's children are the future of tomorrow, and so it's important they have access to nutritious food right from birth through adolescence. So, the POSHAN Abhiyaan (also Known as The National Nutrition Mission (NNM)) was launched by Prime Minister Narendra Modi on March 8, 2018 in Jhunjhunu district in Rajasthan. The POSHAN stands for 'PM's Overarching Scheme for Holistic Nourishment'. The scheme aims to ensure holistic development and adequate nutrition for pregnant women, mothers and children. A key idea of the initiative is to incentivise Anganwadi Workers (AWs) for using ICT-based tools, making it easier to implement and monitor the programme. pregnant women to plan a proper meal for themselves and how to stay disease-free. The POSHAN ABHIYAN launched by the Prime Minister is an initiative to help women and children of our country.

III. IV. IMPORTANCE

Information Communication Technology enabled Real-Time Monitoring (ICT-RTM) is one of the key objectives in POSHAN ABHIYAN. Our website has a 'Dashboard System' for monitoring, supervision and overall user management. The website will be very user-friendly and easy to use for women. There will be different sections for women they will be allowed to borrow a loan at least interest and also they will be trained online via videos lectures on how to nurture their child and take care of them. Vitamin A tablets, food grains and rice, and fortified food packages are available for children and mothers who are showing signs of malnourishment. Weight-for-age growth cards should be maintained for all children under six years of age - children below the age of 3 should be weighed once a month and children aged 3-6 should be weighed quarterly. The women in need of any can contact directly through our website and immediate help would be provided. There will be awareness on how children should be given full vaccinations against six preventable diseases: poliomyelitis, diphtheria, pertussis, tetanus, tuberculosis and measles. Pregnant women should receive a vaccination against tetanus that reduces maternal and neonatal mortality.

IV. BACKGROUND

Malnutrition is the major problem that our country India is facing for many years. In 2018, Our Prime Minister started a scheme called POSHAN ABHIYAAN. The aim of the scheme is to protect children and women of our country with malnutrition, anaemia and stunted growth. Due to an increase in poor people, the number of people malnourished is increasing. So there is a need to promote how we can

help children and

V. OBJECTIVES AND SCOPE

The main objective for developing this project is not only to achieve the targets and objectives set up by the POSHAN ABHIYAAN but also to create the awareness of this programme among the public. The prime focus would be to increase the efficiency of this programme to a very large extent and ensure that it benefits the maximum number of people. The prime objective is to boost the execution of POSHAN ABHIYAAN. The Feb 2020 report shows there are still many states with below-average participation in this programme. Preparation: Based on those results, we will conclude which domains of nutrition need more emphasis and will try to prepare the website on that real.

Designing: We will build the design of our project by pondering over our analysis and accounting to that, we will specify all the necessities required for the creation of the prototype of our project.

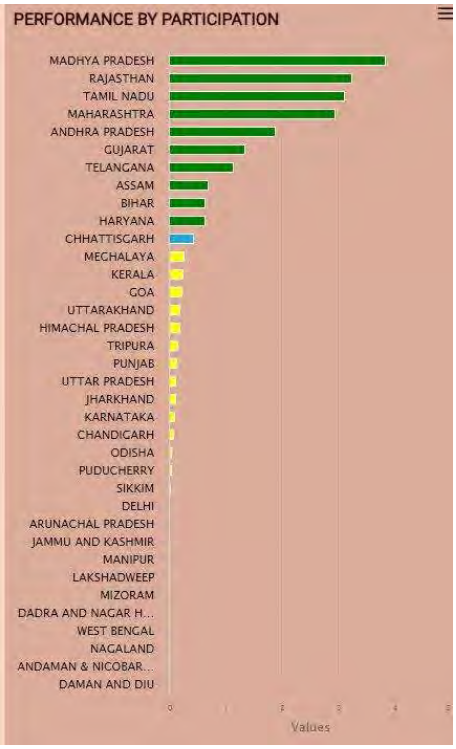
Implementation: The design can now be implemented through web application frameworks like AngularJS and the required databases can be handled through software like SQLite Many states including Uttar Pradesh are still below average in their participation in this programme. We aim to increase participation from these states and maximize the engagement from all parts of our country.

VI. PHASES

Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: We will collect the data about the current status of nutrition among women & children from repositories of government. Based on those we will derive our statistics.

Planning: We will now analyse those statistics to stumble onto some result and will plan accordingly.



Phase 2: Testing and Deployment.

Testing: After the prototype is ready, we will test the UI/UX of our website and will try to ensure that it is feasible with our targeted audience and creates the impact as expected.

Deployment: After completion of integration and testing of the project, the project will be deployed with real-time databases and real-time & operation of the system will be done.

VII. METHODOLOGY

This proposed project will work as an interface between people and government. Users will get to interact with the various programmes of POSHAN ABHIYAAN, the user will also be able to update the information about ongoing programmes in their locality. Users can get the latest updates on POSHAN ABHIYAAN. Users can share their feedback with the government.

VIII. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware: Any Computer System or Smartphone.

Software:IDE:

Intellij IDEA Ultimate 2019.3.1

Front-End:

HTML

CSS

Javascript

Java

Back-End:

MYSQL Database

PHP

IX. PROPOSED UI SCREENSHOT

The above screenshots represent the proposed UI of our website. The Homepage of

the website will contain registration to our website. The user will have to enter his/her

details, and after logging in the user can directly enter his/her problems and they can know more about holistic nutrition. There will be online teaching by professional doctors on how to take care of a child and how to have a proper diet.

IX. OUTPUT

As focus lies on the mother and children as per the National Nutrition Mission (aka NNM or Poshan Abhiyan) by providing guidance and appropriate aid,

the help of website based technology such as these, makes the work efficient. Several families could contact appropriate officials via the website and convey their problems. Since sharing of information has become quicker a better reach to rural areas can be expected. Such technology can be benefitted by the most as The Republic of India has successfully maintained its record of the low-cost Internet, which can be afforded by the middle or lower-middle class, consistently.

X. OUTCOMES

Creating the outline works survey and achievability

study for a proposed solution which will act as an input to the design phase.

Suggesting Statement of work, scope description and scope border for planning the model from the problem definition to decide what needs to be done and what not to be done.

Phase 1 Implementation:

Implementation of the website with a proper database and no failures.

Phase 2 Testing:

Testing was done considering the feasibility of UI of the website among the users and the portability of the website on different devices.

Phase 3 Deployment:

Real-time working was done with different kinds of users from different areas and people were comfortable with our website, they shared good user experience. Thus, the project works efficiently.

XI. RESEARCH GAPS

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. They are as follows: Due to Internet issues in rural areas this website may take time to load on Computers/Mobile Phones. There is a need to provide training to the people in rural areas on how to access/use this website.

XII. SUMMARY

Our website is a platform between the government scheme i.e POSHAN ABHIYAAN and women from various parts of the country. The main focus will be promoting holistic nutrition. It will also contain data from various villages and that data will be submitted to POSHAN ABHIYAAN centre.There will be direct online teaching and camps to spread awareness about the importance of a proper diet and proper food a small child should be given. The website will be made with pure intentions of helping the women and children from various villages and cities through POSHAN ABHIYAAN.

XIII. CONCLUSION

We have successfully shown the implementation of our website. Technologies used to design this poshan abhiyaan website are mentioned. Thus effective execution of the project is done. The data of the people will be reported to the POSHAN ABHIYAAN and it will also be concerned that all people should get the benefits of POSHAN ABHIYAAN.

References

1) NFHS (2017), 4th National Family Health Survey http://rchiips.org/nfhs/factsheet_NFHS-4. shtm
2) Bhutta Z A (2013) Early nutrition and adult outcomes: Pisces of the puzzle Lancet(9891).
3) MoW&CD Government of India: Poshan Abhiyaan
4) NFHS (2017), 4th National family Health survey 2015-16, IIPS, 2017.

Mobile Applications for Police and Citizens Of Crime Records

Ganesh Verma, Yash Pancholi, Rahul Tandel, Ashish Suthar
Thakur college of Engineering and Technology, Mumbai, India
ganeshnv0@gmail.com, yashgreenlawns@gmail.com, rahulkv395@gmail.com, ashishsuthar1000@gmail.com

Abstract: **- In 21st century where mobile and information technology have become an integral part of our lives. A new area where mobile integrated with technology is beneficial for crime reportng since readily accessible information isn't available at any point in investigation this is often a key drawback for communication in police department. Thus, using cloud, we'll attempt to make all the knowledge associated with the criminals available on the Android Application to the police during their investigation which might speed-up the entre process of tracking down the criminals. A mobile application is made available to the common people in order to track down the safest path to reach their destination by giving motivation's Incident was true. This caused chaos among the overall public [1, 2]. The appliance will provide the overall users with the facilities like reportng any incidents which might cause holdup. Moreover, it'll also provide an alternate safe path on user's demand before entering the crime area. The database is going to be stored on cloud to achieve remote access. For avoiding any false incidence to be noted to other user, the data provided are going to be first verified by the police official. These is the intro of our problem statement named is mobile applications for police and citizens for maintaining of crime records.**
The project titled as crime applications for police and citizens is a web-based application. From the app when chosen a crime affected area and also providing an We provide a facility such as reporting online crimes, alternate route.
Keywords -- Authenticated criminal android application

INTRODUCTION

In today's era mobile technology may be employed in many other fields and application like Gaming, Maps, E - mail, Messaging, Photography and then on. One such area is crime area detection and storing criminal data record. A recent mobile application named Mobile Vic PD was released by the Victoria police in Canada for fighting crime. The mobile application may be accustomed report minor crimes, offer anonymous tips to police, stay updated on crimes current, receive missing child reports or check on transferred possession. Because the criminal data isn't available remotely there's a communication gap between the police officials investigating any case. The disadvantage of this application was that it had been liable to fake reportng of crime and there was no other thanks to verify that the Repots and status of any investigations.

PROBLEM DEFINATIONS

In the present scenario fling a FIR/complaint could be a hectic task since one must undergo a awfully huge process for justice. It matters how small or big the crime is. In general, to file a FIR/complaint one must look for a close- by station and must wait until a cop takes upon the request and either we've to recite or must provide it as a written statement. Certainly, this takes nearly every day of an individual's time. During this busy world, it's quite tough to relinquish au courant our routine works. The prevailing system of fling complaints/FIR affects our daily routine, makes us spend plenty of our precious time in it. Fling a FIR, meanwhile we could have done three. Mainly could provide a lot of chances and time for verdicts to simply get away cases making their pre bails ready or perhaps for absconding. To avoid such miserable situations, we've designed an app that creates everything happen in just jiffy.

BACKGROUND

The purpose of this technical is to develop an android app for crime area detection and store criminal records. It provides an application for the user that may provide an alternate path for the users passing by crime area. It allows user to report incidents and procure it verified by the police officials. it'll contains an application for police officials which can perform database operations on record and allows efficient retrieval of required information from the centralized database present on Cloud. The applying targets general public and police officials for managing the incidents and crime without consuming much time. This proposed system is going to be divided into three major modules. This application is going to be useful for the remote access of criminal data which is able to be helpful for the investigations carried by local department. Also, it'll provide the overall users with the facilities like reportng any incidents which might cause tie up. Moreover, it'll also provide an alternate safe path on user's demand before entering the crime area. The database for this project is going to be stored on cloud to achieve remote access. For avoiding any false incidence to be notified to other user, the knowledge provided are going to be first verified by the police officials. After approval of the knowledge it'll be broadcasted to other users using the applying.

Police Application

This module is resulting in the event of police android application which might work as follows. First and foremost, the police have to login with the username and also the password provided to him, as this application isn't publicly available for the final users. After logging into the applying, police are given the features like reporting incidents (crimes and incidents causing traffic jam), view user reported incidents. After the verification of the incident, the database is updated and also the notification is broadcasted to any or all the users who are using this application. Police are given privilege to try to to the criminal database manipulations.

General User Application

This module is resulting in the event of the overall user application which might work as follows. First and foremost, the users will must do one-time registration before using the applying. After registration, user is supplied with the facilities like report incidents, view the notifications and popups which will contain the data like signal and address of the nearby station house, hospital, free station. Moreover, option to view the alternate path are provided by the police officials. User won't run any privilege to form changes within the criminal database. Physical location of the user is tracked with the assistance of GPS which is inbuilt within the mobile phone. To avoid crime affected area, user can request a safest alternate which will be provided by police through the applying.

CloudDatabase

In cloud application, in this paper, cloud is used for storing the database to supply the ability of remote access. As mentioned within the police user application, the username and thus the password employed by the police are visiting be cross-verified with those stored within the database. The complexity of the crime is visiting be selected the first come first serve basis. So as to produce the security to the database SHA-1 algorithm are visiting be used.

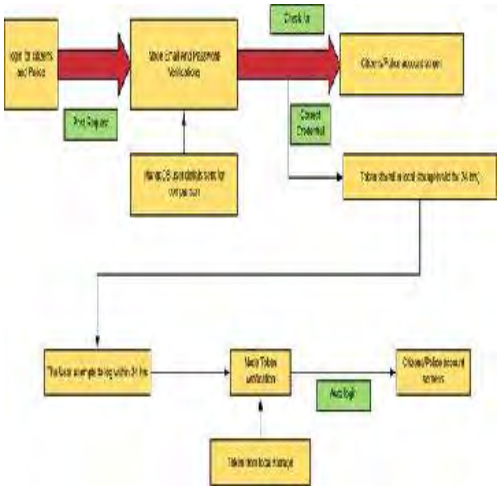
IV.IMPORTANCE

The proposed system consists of three different phases: Installation and configuration of the system, Data Specification of the system, Request/complaint Dispensation, Encryption & Decryption of request using cloud computing. the applying needs to be installed and configured to process the user's request/complaints against crime which is encrypted by the client and thus this information is further decrypted by the respective department of local government. The investigation of the complaint/FIR starts supported the data received from the sender whose details are already per the applying. The method of application allows users to register their complaints against social crimes using their mobile phones. All the complaints registered by public are through their own login credentials which follows by downloading app from their application stores/repositories and signing up into application using phone numbers/email id which generates an OTP to the registered phone number/email id. A brand-new account is created under the registered username in cloud once the OTP accepted by respective cloud.

OBJECTIVES

CREATION OF a mobile application for citizens as well as police to create a sustainable community platform in a society and provide a necessary prevention of crime. The basic idea is to promote safer and better society through crime detection, preventions and control program which include in the program. Through these applications the society will be educated in both ways as an individual and as a group. Crime prevention through mass media is a goal to be achieved in future. This software is made to efficiently and effectively. It results in regular and timely action against crime reported.it can be observed that the information's can be obtained easily and accurately. Assist senior police officer for better management of police force. Reduce manual and redundant records keeping interactions and sharing of information's among police Departments, District, State/headquarters and other police agencies can be made easier and more effective.

METHODOLOGY



TOOLS

Google map (Locations)
Mango DB (secured storage)
Block chain (Securely storing FIR data)
Nodejs-MySQL (Backend)
Lucid chart (Flowchart)
Rest API (fetching data)

RESEARCH GAP

Application pitched by us will help the citizens know about crimes happening around them and lodge a complaint/Fir immediately when they notice a crime happening around us. Verified police accounts will be able to access them and report to the scene as soon as possible. Our app will provide statistics about crime taking place and there solves rate the fir process is very slow and the criminal can get away if action is not taken soon.

EXPECTED OUTCOMES

We can update this system as online application.
The advantage of online application is that, a person can report the crime anytime from anywhere.
Users can view the progress of their complaint online.
The sensitive information can only be viewed by police officers.
The app reduces the time to report crime or for police officers to be able to reach on crime on crime scene on time.
Login for user and admin.

SCOPES

A generic platform for keeping human records from birth till death. Deploying this so platform will serve as a source of information on persons from various states within the country and even those outside. The crime record managements system can then be linked to the present system and thus provide comprehensive data on persons.

The crime record management system should be merged with the criminal justice system to form a mega system that spans over all facets of the justice system. This will enhance synchronization and transfer of data between the court of law and NPF.

CONCLUSION

In this paper we've overcome the matter of communication gap between the police during their investigation. We also provide solution to bridge the communication gap between police and general user. Also, the criminal information will be readily accessible to the police official's as it is stored on the cloud. The problem of reportng fake crimes will be overcome as this application will need the verification of police to report incidents reported by user to broadcast it to other users using the same application. In future, another security algorithm is often wont to provide better security measures for the criminal database. The only challenge of this proposed system is that GPS and the Internet connections has to be activated 24x7. Future research can be dedicated for these challenges.

References

William Akotam Agangiba, Millicent Akotam Agangiba, Mobile solution for Metropolitan Crime Detection and Reporting, Journal of Emerging Trends in Computing and Information sciences, Vol.4, oo. 12, 2013, 2079-8407.

VicPD, Report Crime, Tack Crime, Fight Crime, from your pocket available at: https://www.vicpd.ca/mobile [Accessed: 29/10/2013].

Manav Singhal, Anupam Shukla," Implementation of location based services in Android using GPS and Web Services",(IJCSI) International Journal of Computer Science Issues, Vol. 9, Issue 1, . 2, January 2012, 1694-0814.

Mayur Dhande, Amruta Barawkar, Raman Dhoot," Android Bacha Osos Application", (IJCTA) International Journal of Computer Technology and Application, Vol. 5 (3), 826-828.

Pragya Gupta, Sudha Gupta," Mobile Cloud Computing: The Future of Cloud", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 1, Issue 3, September 2012.

AERO ANALYSTS

Vivek Gupta, Shilpi Char, Suraj Jha, Rinkesh Kumavat, Eshaan Kushwaha, Aniket Jha
Thakur College of Engineering and Technology, Mumbai, India

Abstract- **As technology is spreading its hands in every corner let it be hospitality, infrastructure, communication, education then why travel is returning void output in aspect of technology we know there are many apps, websites available for road, train travel but airplanes and airports are left behind which are the most reliable option for long distance travelling in aspect of comfort.**
Keywords: Arrival Delays, Departure Delays, Visualization, Security, Weather Delays

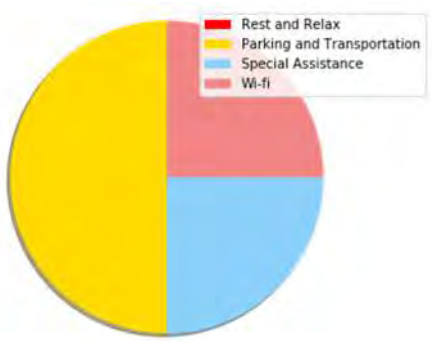
I. INTRODUCTION

"The representation & presentation of data that exploits our visual perception abilities in order to amplify cognition"
– Andy Kirk

Our group has concentrated on this area and is aiming to make a website which will solve general problems related to travelling by air like arrival- delays, departure-delays. It will show which particular airlines are prone to them by default and it is not only limited to delays but also the facilities provided by different airports so that the passengers can get a clear picture what they are going to get through individual airports through visualization which even an illiterate can understand.

II. PROBLEM DEFINITION

A web application is tried to made that is capable of projecting analyzed data through visual tools like bar graphs, pie-charts, packed-bubble on a web dashboard. Airport facilities include rest & relax, special assistance, parking & transportation and wi-fi . the application should have a separate category to visualize delays with respect to airlines and months.



III. C. OBJECTIVES & SCOPES

Classification of dataset based on nature of facility used by a passenger and available facilities provided
Data visualization at pan India level
A centralized dashboard containing visual data representation of facilities, delay of airlines, then delay of months

IV. LITERATURE REVIEW

Modern tourism is impossible without air transportation. The development of airports and tourism are dependent on each other. There are around 80 domestic airports and 12 international airports in India.
During February 2018-Feb 2019, air passenger traffic stood at 316.51 million. By 2020, passenger traffic at Indian airports is expected to increase to 421 million from 308.75 million in 2017-18. Also our country is expected to become third largest aviation market in terms of passengers by 2024.
As millions of people travels through airway, it is the responsibility of airports to give them the best services as possible. The major airports in India where passenger traffic is more are Delhi, Mumbai, Bengaluru, Kolkata, Hyderabad, Chennai, etc. The demand of better services is going to increase every year. Also there are many passengers who are unaware about some of the services that airport offers.
So, there is a need of tool which can rate the airports according to the services they provide and utilization of them by the passengers which will help passengers and airports for their development in our country.

V. METHODOLOGY

This project aims to develop an Application (Aero Analyst) which will be going to visually demonstrate various Airport Facilities including Delay Report, Special Assistance, Air Traffic, etc. This application will act as convenience to understand visual representation of data in a concise manner. The passengers can utilize the available airport facilities better and encourage them to utilize them more. The simplicity of this application is that normal users can understand it by looking at the analytics in the form of bar graphs, pie charts, box plots, scattered plots and histograms and evidently become aware of the features being provided and utilized at the concerned airport.

Facilities

Fig 1.1. implements various facilities of any airport like Rest & Relax, Parking, Wi-fi in the form of pie-chart

Research Design:

Aero Analyst's frontend part is made on Bootstrap, HTML. Visualization of data is done on Python which consists of pandas, Matplotlib libraries. We have also used codeless visualization tool i.e. Tableau which is not an open source tool. The major focus is on visualization of data since it amplifies the cognition abilities of the passengers and maximum amount of data can be monitored easily.

Research procedures:
The data of different airports having different facilities, reviews, delays, etc. are

collected from major websites. Collected data of around 20 airports at pan India level and state level. The process of visualization consists of different paths Data Collection, Data Cleaning,

Data Prediction. The raw data collected was in unstructured manner then later it was converted to structured data which helps us demonstrate the facilities of the airports.

The visualization will provide answers to these questions:
Which City has the most traffic?
Which city has the least?
Which month is the busiest in the year?
Which airport route is the busiest one?
Which airports have special security management for the passengers?

Selection and access:
The application design consists of major three points i.e. facilities, delays, air traffic. These analyses can be used by the airport authorities for better management of utilities and push forward better airport advertise. Passenger can efficiently manage their time at the airport by knowing the various that are provided by the airport and conveniently use them.

The airport companies can easily demonstrate the monthly and annual delays of their airplanes and can come to a solution in future in
advancement of their techniques. Insightful information in finger tips will make passengers, staff to take the right decision. The dashboard of aero analysis also consists visualization through maps which gives end to end management of security of passengers.

Most of the money on airports are used in better infrastructure and security surveillance of the airports. Passengers invests around 4 hours for security and baggage check in on airports .The workload is reduced by using the dashboard as it contains a special security visualizations that is connected to the airport monitor .Every passenger is rated according to their previous criminal records. The visualization helps to maintain airports secured as shady peoples will be colored" red "on to the scatter plot and normal citizens are colored "blue". The scatter plot will help to secure the airport from inside which reduces the chance of hijacking.

FUTURE ADDONS:
Analytics on passenger data
Cleanliness and efficiency of machinery at airports
Air traffic density and route analysis

BENEFITS FOR THE SOCIETY:
People can easily compare which aviation provides them the fastest arrival to their destination.
It will make the airport authority to easily
enhance their facilities according the users feedback.
Map visualization of all collectively routes
will help the passenger as well as airport authority to predict the future air traffic.

VI. HARDWARE & SOFTWARE REQUIRMENTS

Hardware: This website can be access from any Operating System like Windows, Android, Linux, MacOS, etc.
Software:

ANDROID STUDIO FLUTTER, REACT NATIVE

Front-End:
HTML
CSS, BOOTSTRAP 4
JavaScript

Back-End:
MYSQL Database
PHP
PYTHON with DJANGO

VII. STATEMENT OF LIMITATIONS

There are many booking websites which displays the data like timings, delays etc. But none of them provides information about the services they provide. Till now we don't have such application which can rate airports on the basis of services they provide.

Our research will be beneficial for airports as well as

people who travels through airway. It will help airports to know that where they are lagging in providing services to the passengers and can modify them accordingly. Also it will help passengers to utilize maximum available services at the airport.

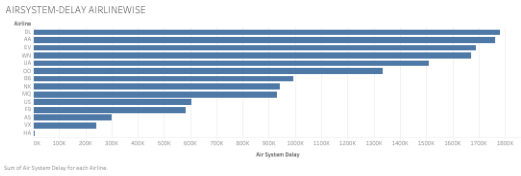


Fig 1.2

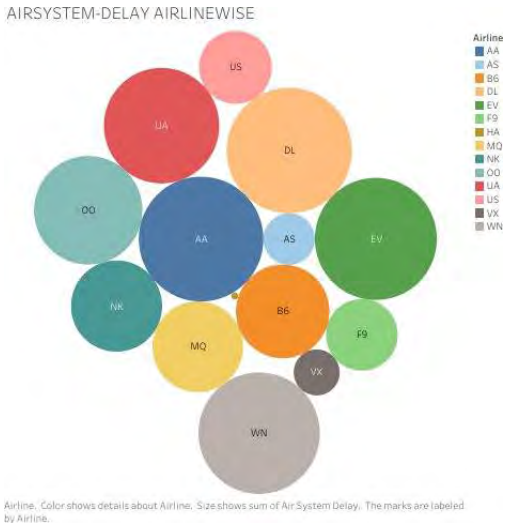


Fig 1.3

Both Fig shows the air-line delay but Fig 1.2 is represented in the form of bar-graph and Fig 1.3 is represented in the form of bubble graph.

VIII. RESEARCH GAPS

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too.
They are as follows:
This website may take more time to load at several places due to less Internet bandwidth
The real-time data may fluctuate as due to the geographical region.

IX. CONCLUSION

The following project helps in the clear visualization of data provided by the concerned associates, which is processed by the algorithms created by our programmers to create extremely easy to understand analytics in the form of graphs and plot. These analytics can be understood by most individuals.

Proper utilization of large data is extremely essential and should be carried out in an efficient manner. The procedures carried out during the project, helps in minimizing complications and ambiguity by reducing it to a visual and comparative study.

The website is beneficial to both passengers as well airport authorities. Passengers can refer it to plan out their trips more efficiently and also become aware of the various facilities being provide at a concerned airport, it promotes higher usage of the facilities as well.

References

1) Wikipedia, Airports authority of India https://en.wikipedia.org/wiki/Airports_Authority_o f_India
2) Aviation-mar-2019
3) https://www.ibef.org/download/aviation-mar- 2019.pdf
4) Govt. of India
5) https://data.gov.in/

Real-Time Based Facial Recognition Systems.

Jyoti Khare, Aman khairwar, Abhishek Tripathi, Hinal deriya, Mohammad Ali khan
Information Technology Thakur College of Engineering Kandivali(E), Mumbai-400101
jyotirkhare@gmail.com,khairwaraman@gmail.com,Possible.abhishek90@gmail.com, hinalderiya03@gmail.com, sameerjoshi67@gmail.com, Khan907666@gmail.com

Abstract—**Face recognition is used to prove identity across a wide variety of settings. Despite this, research consistently shows that people are typically rather poor at matching faces to photos but, when you work with machine this introduces a new level of identification through a real-time algorithm. How to accurately and effectively identify people has always been an interesting topic, both in research and in industry. With the rapid development of artificial intelligence in recent years, facial recognition gains more and more attention. Compared with the traditional card recognition, fingerprint recognition and iris recognition, face recognition has many advantages, including limit to non-contact, high concurrency, and user friendly However, face recognition skills are subject to wide individual variation, with real time machine showing exceptional ability—a group that has come to be known as 'Bureau of Police Research & Development". The Bureau of Police Research & Development adapt 'Real-Time Based Facial Recognition Systems' from within its ranks, for deployment on various identification tasks. Here we test Facial Recognition Systems from within this Bureau, and ask whether they are really able to perform at levels above control groups. We consistently find that this 'Facial Recognition Systems' perform at well above normal levels on tests of unfamiliar and familiar face matching, with degraded as well as high quality images. Allocating this system not only remove difficulties but also include more automation and efficient working environment.**

Keywords—Deep neural network, face recognition, server-client model, deep multi-model fusion, convolutional neural network.

I. INTRODUCTION:

"Faceprint are one way that technology recognizes individuals"

As the digital world and real-world merge more and more together, how to accurately and effectively identify users and improve information security has become an important research topic. Not only in the civil area, in particular, since the 9-11 terrorist attacks, governments all over the world have made urgent demands on this issue, prompting the development of emerging identification methods. Traditional identity recognition technology mainly relies on the individual's own memory (password, username, etc.) or foreign objects (ID card, key, etc.). However, whether by virtue of foreign objects or their own memory, there are serious security risks. It is not only difficult to regain the original identity material, but also the identity information is easily acquired by others if the identification items that prove their identity are stolen or forgotten. As a result, if the identity is impersonated by others, then there will be serious consequences. Different from the traditional identity recognition technology, biometrics is the use of the inherent characteristics of the body for identification, such as fingerprints, irises, face and so on. 6 Compared with the traditional identity recognition technology, biological features have many advantages, as:1. Reproducibility, biological characteristics are born with, cannot be changed, so it is impossible to copy other people's biological characteristics. 2. Availability, biological features as part of the human body, readily available, and will never be forgotten. 3. Easy to use. Many biological characteristics will not require individuals to corporate with the examine device. Based on the above advantages, biometrics has attracted the attention of major corporations and research institutes and has successfully replaced traditional recognition technologies in many fields. And with the rapid development of computer and artificial intelligence, biometrics technology is easy to cooperate with computers and networks to realize automation management, and is rapidly integrating into people's daily life. When comparing the differences between different biometrics, we can see that the cost of facial recognition is low, the acceptance from user is easy, and the acquisition of information is easy. Facial recognition is the use of computer vision technology and related algorithms, from the pictures or videos to find faces, and then analysis of the identity. In addition, further analysis of the acquired face, may conduct some additional attributes of the individual, such as gender, age, emotion, and et

II. PROBLEM DEFINITION

A complete face recognition system includes face detection, face pre-processing and face recognition processes. Therefore, it is necessary to extract the face region from the face detection process and separate the face from the background pattern, which provides the basis for the subsequent extraction of the face difference features. After detection ,the details will be searched cross various websites and social medias to gain access to information regarding the image. This will make the police bureau easy access to criminal's personal information if available on any of the websites.

III. BACKGROUND:

Pioneers of automated face recognition include Woody Bledsoe, Helen Chan Wolf, and Charles Bisson. This project was labeled man-machine because the human extracted the coordinates of a set of features from the photographs, which were then used by the computer for recognition. Using a graphics tablet (GRAFACON or RAND TABLET), the operator would extract the coordinates of features such as the center of pupils, the inside corner of eyes, the outside corner of eyes, point of widows peak, and so on. From these

coordinates, a list of 20 distances, such as the width of mouth and width of eyes, pupil to pupil, were computed. These operators could process about 40 pictures an hour. When building the database, the name of the person in the photograph was associated with the list of computed distances and stored in the computer. In the recognition phase, the set of distances was compared with the corresponding distance for each photograph, yielding a distance between the photograph and the database record. The closest records are returned. Some face recognition algorithms identify facial features by extracting landmarks, or features, from an image of the subject's face. For example, an algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. These features are then used to search for other images with matching features.

Other algorithms normalize a gallery of face images and then compress the face data, only saving the data in the image that is useful for face recognition. A probe image is then compared with the face data. One of the earliest successful systems is based on template matching techniques applied to a set of salient facial features, providing a sort of compressed face representation.

IV. IMPORTANCE

When you post a photo on Facebook, and the platform automatically tags the people in the image, you might not give much thought to the technology behind the convenience. However, when you discover that facial recognition technology could track you without your permission while you walk down a street in Mumbai, it might make you question the invasion of your privacy. Just like with any other new technology, facial recognition brings positives and negatives with it. Since it's here to stay and expanding, it's good to be aware of the pros and cons of facial recognition.

Cons of facial recognition: 1. The technology isn't as effective at identifying people of colour and women as it is white males. One reason for this is the data set the algorithms are trained on is not as robust for people of colour and women. Until this is rectified, there are concerns about the ramifications

for misidentifying people with the technology.

2. In order to benefit from the positive aspects of facial recognition, our society is going to have to work through some significant challenges to our privacy and civil liberties. Will individuals accept the invasion of their privacy as a proper cost to being more secure and for the conveniences facial recognition provides?

V. OBJECTIVES AND SCOPE:

1. To track the full face during its motion in a video.

2. To track many faces during conversations.

VI. 3-DIMENSIONAL RECOGNITION A newly emerging trend, claimed to achieve improved accuracies, is three- dimensional face recognition. This technique uses 3D sensors to capture information about the shape of a face. This information is then used to identify distinctive features on the surface of a face, such as the contour of the eye sockets, nose, and chin. One advantage of 3D facial recognition is that it is not affected by changes in lighting like other techniques. It can also identify a face from a range of viewing angles, including a profile view. Three-dimensional data points from a face vastly improve the precision of facial recognition. 3D research is enhanced by the development of sophisticated sensors that do a better job of capturing 3D face imagery. The sensors work by projecting structured light onto the face. Up to a dozen or more of these image sensors can be placed on the same CMOS chip— each sensor captures a different part of the spectrum. Even a perfect 3D matching technique could be sensitive to expressions. For that goal a group at the technique applied tools from metric geometry to treat expressions as isometries a company called Vision Access created a firm solution for 3D facial recognition. The company was later acquired by the biometric access company Bioscrypt Inc. which developed a version known as 3D FastPass.

Vehicle Recognition And Compilation Of Database Software

Abhishek Pandita, Deeksha Rai, Aditi Singh
TCET, MUMBAI
abshek2109@gmail.com, raideeksha809@gmail.com, singhaditi206@gmail.com

Abstract— **Today's age is an age of AI, ML and IOT and it is a digital era where everything is going to be automatic i.e. less need of human work or you can also say no need of humans to do work at all . Most of the human work is done by Hardware machines and Software systems.**
Everything is done in just one click. The only work of humans is to just wait and watch. The same is needed to be done in department/field of Vehicle Recognition and the applications related to it. Taking this into consideration the Vehicle Recognition Software is to be developed which will help in many ways like it will be helpful in parking areas, toll booth, maintaining law and Supressing criminal activities. The System will help in multiple fields and also the work load to humans will be decreased. This system will also help in better functioning of law and desired work. The system will consist of a software and hardware mainly.
Keywords – Vehicle recognition; Number plate; Compilation; RTO, ANPR-Automatic Number Plate Recognition, OpenFace.

I. INTRODUCTION

Vehicle Recognition and Compilation Of Database
is a software that recognises vehicle number from number plate and the face of driver. This technology consists of both software and hardware. Hardware mainly comprises the camera or detector that captures the face of driver and number plate of vehicle. Software will extract the number form number plate and stores it in database, the captured face will also be stored in database where it will be matched with the bio data of people for extracting the information of driver.

This software will help in many ways like it will help to know if the vehicle is over-speeding , if the traffic rules are broken by the driver and the RTO wants to check the registration of vehicle , if the vehicle is overloaded, the software will also help in case of theft, in case of toll on highways .The camera is fast enough that it can capture the number plate of fast moving vehicles.
The database of driver will be created when the vehicle is registered in RTO. The software has a special type of technology that converts the number plate image into text. The number plate should be according to the rules of RTO Terms and Condition. The camera will also have night vision, so that it can capture the images of vehicle in day as well as night.

II. PROBLEM DEFINITION

Design a software for vehicle recognition and compilation of database. The software should get input form hardware i.e. camera and read the image and match the information with database information and to do the task assigned (depends where it is used).The software should be useful in case of law enforcement ,toll booth, parking system, RTO offices etc. The software should use modern Algorithms and searching techniques for fast and accurate results.

III. BACKGROUND

Vehicle recognition is a deep learning based software that is available as computer software. Unlike other license plate recognition software's it has multiple
uses and has an additional face recognition technique which will help in case of theft and other crimes.

The license plate recognition technique algorithm first debuted in 2007 but the ANPR technique was invented in 1976 at police scientific development branch .Since then it is been developed and many new algorithms are introduced and proposed but no system is efficient for proper functioning of the system and the work assigned to it.
The Vehicle recognition uses the best and latest Algorithms for both facial recognition and number plate recognition for best and efficient functioning of the software. This is the best type of intelligent transport system. Previous methods have many names like ALPR,CPR,LPR,VRI and much more but no technique is efficient till now.

IV. IMPORTANCE

Vehicle Recognition has wide range of applications as number plate of vehicles is primary and mandatory identifier for vehicles. Vehicle Recognition helps in automatic content access and management of database of vehicle and drivers information. Some of its important applications are below--

Parking--The vehicle Number plate is recognised while entering in parking lot and the bill will be generated on account of driver, identified by facial recognition.

Traffic Control--The overloaded vehicles will be charged fine and the vehicles which are over speeding or breaking traffic rules will also get charged with challans.

Law Maintenance--Theft of cars from parking lots or garages can be prevented when the vehicle number plate and drivers face is recognised by software. The software will automatically send message to the owner of car if the car is spotted with other driver, also the car can be tracked by other cameras.

Border control--The unidentified vehicles can be spotted on borders and the drivers face can also be identified to alarm the guard's.

V. OBJECTIVES AND SCOPE

The main objective of building the system was to build an effective software and hardware that can help in recognising the vehicle by number plate and drivers face. The software is useful in Traffic control, Parking System ,Law maintenance and Border control.
The software can work efficiently in both day and night modes. The additional feature of the system is to recognise both face ID and number plate at the same time. It can also be used on entry gates of the buildings and apartments or college campus by recognising the face ID of driver or number plate of vehicle. Only the recognized and registered vehicles will be given entry.

VI. PHASES

Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: First work to do is to get strong knowledge on this project title and get description preparation on it. After doing survey, some research is needed to be done for better understanding of topic in depth.

Designing: After research and survey, demo model/design will be built so as to read the prototype of project and then design the main project model. This will also help in checking the drawbacks in project.

Phase 2: Testing and Deployment.

Testing: After the model is ready the project will be tested on small amount of data (In a Building or Campus) to check the proper functioning of the system. If any issues are found then the issues will be resolved in the model itself and the model will be tested again.

Deployment: After testing and finalizing the model, the system will be deployed on the main one and the real time testing will be done.

VII. METHODOLOGY

The proposed system will automatically detect the vehicles and take snapshot of their number plate and driver and send the data to the System's software part for further work. The software will extract the text from the number plate and the face ID from the images and check for the information in the database obtained from RTO.

VIII. HARDWARE AND SOFTWARE REQUIREMENTS

- Resolution: 1080x720 and above, up to 1080p
- Alarm Center Workstation: Core i9, 12GB RAM, 2TB HDD Windows 10 OS
- Frame Rate: 20 fps and above
- Database: MySQL

- 1 HDMI to VGA converter
- 1 USB cable

IX. XI. OUTPUT

When the software gets the image from cameras the system will start generating information data from that image and start checking for the same matching data in database.

X. BLOCK DIAGRAM

Fig1. Block Diagram[4]

The above block diagram represents the working of the project. Here the input image is given to the software which is processed using the Algorithms.

XI. CIRCUIT DIAGRAM

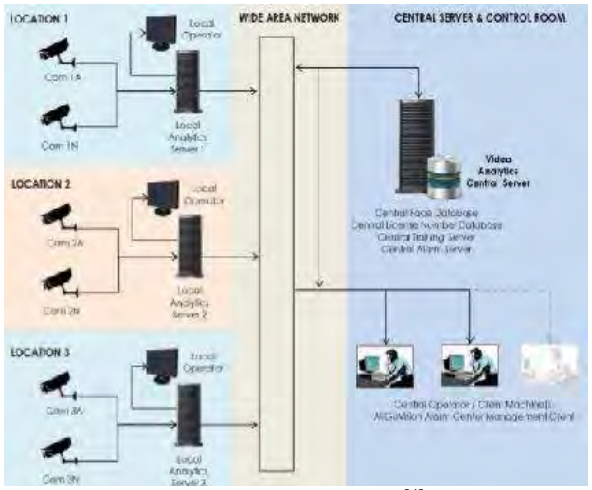


Fig2. Circuit Diagram[4]

This circuit diagram represents the internal working of the system. It consists of Hardware Part i.e. camera, the software part and the server database.

XII. RESULT

The software will match the information collected by the cameras and the information in database and do the useful work assigned to it as per use.

Fig3.Result and Output Diagram[4]

XIII. OUTCOMES

Creating the outline, survey work and research study for the proposed problem statement which will help in design phase of the project.

The work will help in suggesting that what work is needed to be done and what is not.

The research work was done to check the drawbacks of previously proposed systems and what is needed to be done further. Also, the survey and research helped to choose which algorithm is best for the work.

According to Survey work the Segmentation and Extraction algorithm is best for number plate recognition and for face recognition OpenFace algorithm is best.

XIV. RESEARCH GAPS

No project is 100% perfect. This project also has some [4] https://www.allgovision.com/index.php drawbacks. They have been discussed them below:

The system fails to identify the Face ID of drivers who's over speeding or if the vehicles windshield is not clean.

The system will not identify the number plate of vehicle if the number plate is not same as the guidelines provided by RTO office.

The system will fail if the Vehicle will be having a large number of printed characters or numbers on it. The system will get confused.

XV. SUMMARY

The vehicle recognition system software focuses mainly on detecting the number plate area of a vehicle and the windshield of a vehicle then takes pictures of it and sends to the software part of system. There the software uses algorithms for extracting the data from the images and matches it with the data in database which is created while registering the vehicle. After matching the data, the software will do the required and desired task assigned to it. For reference this paper also has an abstract and introduction at beginning. The paper also has importance, objectives and scope outcomes in it.

XVI. CONCLUSION

The Vehicle Recognition and Compilation of Database Software is a system software which uses two modern and efficient Algorithms for number plate recognition and face ID detection and then performs the desired task. It has many applications and future scope. The weakness of the system is also discussed in this paper.

References

[1] Hendry, Rung-Ghing Chen, "Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning", Image and Vision Computing, Volume 87, Page 47-56.

[2] Yousri Kessentini, Mohamed Dhia Besbes, Sourour Ammar, Achraf Chabbouh, "A two-stage deep neuralnetwork Recognition.", Ecpert Systems with Applications, Volume 136, Pages 159-170.

[3] **Yu Wang, Xiaojuan Ban, Jie Chen, Bo Hu, Xing Yang,** "License plate recognition based on SIFT feature",Optik, Volume 126,Pages 2895-2901.

Efficient Chatbot Designing

Akshat Chandel, Darshil Ajudia, Parth Desai, Hinal Kuvadiya, Pratik Bhatt, Manav Agrawal
TCET,MUMBAI
akshatchandel92@gmail.com, darshil1999@gmail.com, parthdesai0849@gmail.com, hinalkuvadiya@gmail.com, pratikbhatt503@gmail.com, manavagrawal12@gmail.com

Abstract— **It is very important to know what a chatbot is. So, a chatbot is a software or an interface to get work done via text messages or voice command-based conversation with computer program which may or may not be aided by a human.**
It can be referred as a conversation of a human with robot. It's very important for today's media and industry to integrate chatbots in their environment and so it's very important that the chatbots should be designed accordingly.
This paper deals with how to design an efficient chatbot that can be used a base template by any user and set some ground rules that should be followed by every user designing a chatbot so that it can be interactive like a human and can even provide responses based on human like thinking.
Basically, we would be emphasising on what chatbots really are and how are they so important in this era and also how to use them and how to create them.
Keywords –, chatbot; efficiency; conversational interactions;

I. INTRODUCTION

"Chatbots represent a whole new trend in how people access information, make decisions, and communicate"
This paper deals with what chatbots are, what they can be used for and how to develop an efficient chatbot. Most of the common people don't know what chatbots are and what potential it bears in todays technology age, so really what is a chatbot? "A robot" i.e. an autonomous program that can

imitate, behave and respond like a human being. So, in simple words CHATBOT can be defined as an automated computer program that can respond to the users in a human way and solve their queries and doubts in real time.
Designing an efficient chatbot requires some thought process about the use and needs of the user. So, there are some steps and rules to be followed which will be discussed later in this paper in designing a chatbot. Designing an efficient chatbot is very significant because it needs to meet the expectations of the company and the customers. Also, most users don't want to waste their time by speaking to a bot rather they prefer to talk a human and so designing an interactive and human-like bot is very vital. According to a survey, 60% of the users have claimed that a chatbot has failed to resolve their issues, so this may be very crucial for any company or an entity as it may lose its potential customers due to their unresolved issues and doubts. A chat bot needs a good script, name, platform and an image for it to work.

II. PROBLEM DEFINITION

Many chatbots are nothing more than glorified flowcharts, their responses fumbling forth from rigid IF/THEN scripts. Even artificially intelligent chatbots, though skilled at detecting patterns in human language, are lacking when it comes to natural language understanding. Inevitably, chatbots that draw replies from IF/THEN scripts will run into a question or request that wasn't accounted for. When this happens, most bots will attempt to recover by asking a clarifying question that redirects conversation back to the safety of their predetermined responses, but problems arise when a bot's corrective questioning leads to a conversational dead-end or places blame on the user, even subtly. Chatbots run the risk of being impersonal if designers aren't keenly aware of their user's practical and emotional needs when seeking help from a bot. Some chatbot designers are attempting to offset the impersonal nature of bot interactions by mining more and more user information. It seems obvious that this is a bad idea, but some companies continue to learn the hard way.

III. BACKGROUND

Chatbot also known as Chatterbots or chatter robots, is the computer system that can communicate with human in the form of messaging app. They can understand multiple question requested by humans. They also have the ability to differentiate between uniqueness of words including emoticons. In order to get the best quality of chatbot conversation, they need to have richness of vocabulary amongst people. Chatbot may look like a normal
messaging app, they have the application layer, a database and also APIs(Application Programming Interface) working at the background . User interface represent the interface to make easy contact with user. While Chatbot is easy to use, at background it has the complexity to achieve. Most of the chatbots have log of the conversation and the developers use these logs in order to understand user requests. The log is then used to improve Chatbot conversation. Chatbot works by matching the questions from user with the help of machine learning. The developer needs to train the chatbot to understand different questions and to deliver the suitable output. Chatbot is being trained through the analysis of thousands of logs from human conversations. More the number of logs the smarter the chatbot will be.

IV. IMPORTANCE

Chatbot applications streamline interactions between people and services, enhancing customer experience. At the same time, they also offer companies new opportunities to improve the customer engagement process. They also yield better operational efficiency by reducing the typical cost of customer service. To be successful, a chatbot solution should be able to effectively perform both of these tasks.

Human support plays a key role here: Regardless of the kind of approach and the platform, human intervention is crucial in configuring, training and optimizing the chatbot system. Even though they are increasingly used, the modern chatbot is still a young technology. With the continuing development of AI, the potential for bots in business and personal lives is unlimited. They can be easy to build. It depends on what you want to achieve.

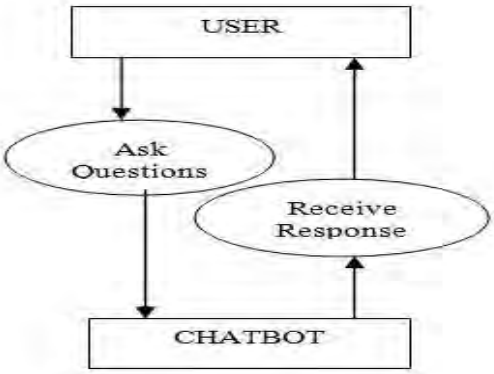
V. OBJECTIVES AND SCOPE

1. The amount of information present on a website is not accessible to an average individual if the visitor cannot find it. Chatbot can helps to make information accessible to all.
2.There is sharp decline in people downloading apps as some app which are not in too much use take to much storage. Chatbots will help get more customers and people who arrive at the website are able to ask queries and be answered more efficiently and quickly.
3.Chatbots can help to schedule meetings and broadcast newsletter.
4.Multilingual chatbots can communicate with people their comfortable in and attend their queries more efficiently.
5.Humans who can only communicate with one person at a time but chatbots can communicate with thousands of people simultaneously.
6. Chatbots are virtually robots hence they do not get tired and because of which they can work 24/7 and it will help to increase people to visit the website more since they get quick replies to their issues.
7.Chatbots helps to save money and are easy to configure for people having different needs and use of it.The cost of keeping them updated is very low.

VI. DESIGN OF CHATBOT

"Designing a bot is like brewing beer; it always takes the same ingredients to make it, but slight changes to the process can transform the whole outcome."

A chatbot is a computer program used to simulate real world communication. The chatbot is implemented to enact a conversation with user. The chatbot effectively responds to question posed by the user. Chatbots are not new invention, history of chatbots can be found during 1950–60s when Alan Turing & Joseph Weizenbaum contemplated the concept of computers communicating like humans do with experiments like the Turing Test and the invention of the first chatterbot program, Eliza. It helps the user by answering the questions asked by them. The program is implemented using Java programming language. Particularly Java applets are used. Applets are used because it is easy to create the dialog box required for the conversation between the user and the bot. The design of a chatbot is represented using the following diagram



The following points must be considered while designing an efficient Chatbot –

[A] *Selection of Platform*

The usage of a chatbot cannot be restricted to a single platform, hence while designing it the developer must really emphasis on making it platform independent.

[B]*Selection of Software*

Java is primarily the most used language to write chatbot for various needs. However, other programming languages such as Python, Closure, PHP, Ruby are also widely used to create robust chatbots.

[C]*Creating a Chat*

The chat is created using a pattern that is known to the user and could be easy to understand. Chat dialog box show up to create conversation. This dialog box is created using java applets.

[D] *Pattern Matching*

This process makes use of Artificial Intelligence to help generate response text for the chatbot. The input provided by user is saved into the database and is subsequently answered with an appropriate response.

[E]*Human Personification*

The way a chatbot responds to a user is hugely responsible for the growth of business or institution where it is being used. The conversation with the customer must reflect that a conversation with a real person is taking rather than computer program. The chatbot should be made witty and entertaining to ensure maximum user satisfaction.

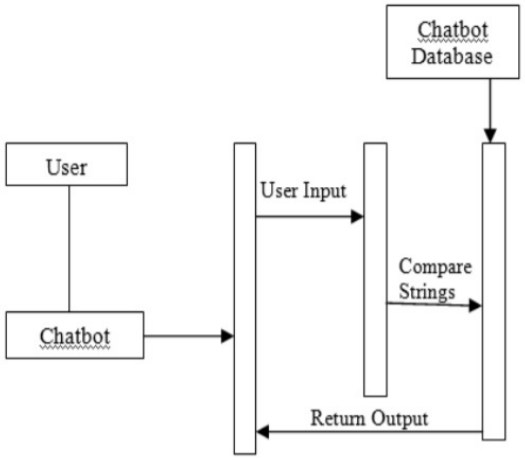
VII. IMPLEMENTATION PROCESS

[A]*Creating the dialog box on display*

All the essential packages required for creating the dialog box are imported beforehand. The size of the dialog box and text area inside the dialog box visible to the user is adjusted and implemented as per the rquirement. Vertical scrollbar is used so that the screen gets scrolled up as the conversation goes on. Horizontal scrollbar is avoided because the size of the dialog box is fixed.

[B] *Creating a database*

2-D string arrays are used for the creation of database. Rows in the array are used for request and response. All the even rows contain the request or questions and all the odd rows contain the response or answers. Columns in the array are applied to save different types of questions that could asked by the user and responses that a Chatbot can answer. There is one row in the array which contains default responses which is used when the matching question is not found in the array.



VIII. SOFTWARE REQUIREMENT

There are various tools that can be used to create a chatbot IBM Watson, Microsoft Azure Bot Service, QnA Maker, Semantic Machines, Recast AI. Once the chatbots are created we need to deploy them, the deployment of a chatbot depends on the user requirement. Some common platforms for deployment of a chatbot are Facebook Messenger, Slack, Skype for business, Facebook Workplace, Kik.

IX. IX. RESULT

Most people are attracted to a system that is human-like. And many users do not know that Chatbot will not only give feedback in the form of text and voice command whereas Chatbot nowadays have an interactive way of serving information using graphical interaction or graphical widget. Th main benefit of using chatbot is it is able to reach broad audience even from great distance only using the messenger apps. Beside that, this automated human-computer conversational platforms works positively to provide efficient service in various field to serve human in many ways. In this paper, the review is based on the design work, features, how it interacts with users and also its interface. Chatbots are actually an innovative approach to automate user-personalized messages. If the chatbots are well designed and implemented, it could be a tool to attract user engagement and provide good user experience between humans and the served field. The development of chatbots should carefully planned, choosing the appropriate platform tools is very important since it can help in boosting the effectiveness and efficiency of the chatbots.

X. X.CONCLUSION

We have successfully shown the importance of chatbot and its uses. It is reconfigurable in any sector and it can also be used for multiple task. Designing of efficient chatbots will meet the expectations of companies and customer. A chatbot i.e. designed humanlike bot will help customers to interact with them properly and get their queries solved. A multilingual hatbot will help users to ask queries in their preferred language and it help them to be comfortable while asking. Efficient chatbot will help in rise in business. As they are cheaper than hiring an employee company can buy them easily and chatbots won't take much time to work according to the company needs whereas employees can even take months. This will also help in making meetings, appointments, and also broadcast newsletters. This will also minimize the workload on humans. Chatbots require relatively low update cost, so it will help to attend more customers than average. This knowledge should help exploring further possibilities of chatbot and it should enable more companies to apply chatbots to new works and thereby also improve human-machine interaction.

Refrences

[1]. https://chatbotslife.com
[2]. https://www.geeksforgeeks.org
[3]. https://en.m.wikipedia.org/wiki0
[4]. https://www.userlike.com
[5]. https://chatbotsmagazine.com
[6]. https://www.lollypop.design

App for Online Opd Appointment & Hospital Information System

Ritika Chaube, Rahul Kurmi, Richa Gupta, Neeraj Chauhan, Nilesh Gond, Chaithanya Moozhickal
Thakur College of Engineering and Technology, Mumbai, India
ritikarchaube@gmail.com, kurmirahul557@gmail.com, richag536@gmail.com, chauhanneeraj170@gmail.com, gondnilesh123@gmail.com. binuchaithanya0015@gmail.com

Abstract- **Appointment scheduling systems are used to manage access to service providers. Many factors affect the performance of appointment systems which include arrival and service time. Thus a proper scheduling system has to develop by considering all these factors which will increase patient satisfaction, which in turn increases profit. An online scheduling system allows individuals to conveniently and securely book their appointments online. Compared to the usual queuing method, the web-based appointment system could significantly increase patient's satisfaction with registration and reduce total waiting time effectively.**

Keywords: Online Appointment System, Scheduling

I. INTRODUCTION

The term of "appointment" refer to the period of time allocated in the schedule to a particular patient's visit and "service time" refer to the amount of time the physician actually spends with the patient. Appointment scheduling can be classified into two broad categories. Static and Dynamic. In static appointment scheduling all decisions must be made prior to the beginning of a session, which is the most common appointment system in healthcare. In dynamic appointment scheduling the schedule of future arrivals are revised continuously over the course of the day based on the current state of the system.

The appointment-scheduling process which is either dynamic or static can be viewed as burden in hospitals, which can be eliminated through an efficient online appointment scheduling system. The benefits of implementing this technology touch everyone involved in the scheduling process. Administrators and staff can conduct their tasks more efficiently and accurately, while patients have the ability to book their appointments and reservations quickly and more conveniently Patient can also provide additional information to the doctor, making the doctor aware of their situation and giving the doctor time to prepare the necessary information for when the patient's arrives. For doctors, online appointment scheduling brings a lot of value add services and being able to store patients' data securely for future reference. But the most wonderful and useful advantage is that online appointment scheduling is amazingly low cost.

II PROBLEM DEFINITION

The appointment-scheduling process which is either dynamic or static can be viewed as burden in hospitals, which can be eliminated through an efficient online appointment scheduling system. The benefits of implementing this technology touch everyone involved in the scheduling process. Administrators and staff can conduct their tasks more efficiently and accurately, while patients have the ability to book their appointments and reservations quickly and more conveniently. Producing standard appointments of 15, 30 and 45 minute durations. In contrast, paediatric practices, and some family practice groups, tend to use 10 minute increments with resulting planned visits of 10, 20, 30 and 40 minutes. In this method physicians can reduce down time and the need for double booking by calibrating scheduled time closer to actual visit time. But it increases the complexity for scheduling personnel.

III BACKGROUND

A number of different methods are available to schedule appointments in the medical office. They include the following:

- Double Booking
- Ten Minute Increments
- Modified Wave Scheduling.
- Staggered Starts
- Group Meetings

IV IMPORTANCE

Modified Wave Scheduling: In a modified wave system, patient appointments overlap so that when one finishes early, another patient is waiting to be seen by the physician.

Staggered Starts: If a modified wave schedule is not appropriate, some of the same benefits can be achieved by staggering visits in five or 10 minute intervals.

Group Meetings: By seeing such patients as a group, some physicians have found they can save time, create a highly supportive atmosphere, and devote more time to patient education and health issues than would be possible during traditional office visits

Double booking: Double booking appointments is a basic technique for minimizing down time by ensuring that there always is a patient ready to be seen when the physician is available

Ten Minute Increments: Internists traditionally design schedules around 15 minute time increments, thus In traditional appointment system patients has to come to the hospital and queue at the appointment window to make the appointment. But they usually end up waiting for very long periods of time. The patient can, however decide to schedule an appointment, but this option does not usually work well for all parties

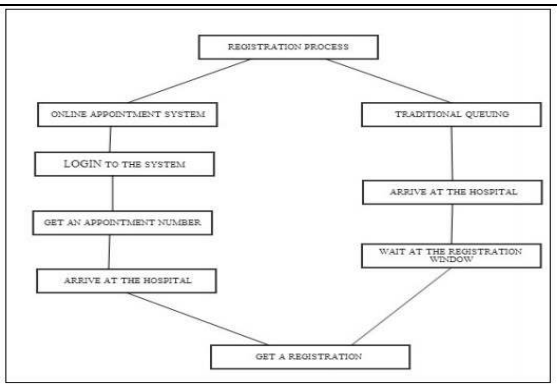
involved. Online appointment System provides the following benefits:-

- Schedule patients for various medical procedures-test, treatment
- View daily, weekly, monthly patient schedule
- Create Patient Records and Appointment reports
- Helps to trace Patient Flow based on arrival, visit and departure time
- Avoid no-show, missing, over- booking patients and other conflicts
- Email appointment schedule reminder
- Provide daily appointment scheduling reports.The flexibility of online scheduling system enables it to be utilized for a variety of different services and activities such as:
- All the appointment can be aggregated from the website, phone enquiries or direct enquiry
- Scheduling patient appointments, treatments and services
- Booking vaccine

V.OBJECTIVES AND SCOPE

The main objective of developing this project was to build a online OPD appointment & hospital informative system. In traditional appointment system patients has to come to the hospital and queue at the appointment window to make the appointment. But they usually end up waiting for very long periods of time. The patient can, however decide to schedule an appointment, but this option does not usually work well for all parties involved. Parties involved includes: the patient, the medical personnel and the hospital. The patient wishes for readily available and convenient appointment times. When they do not find a close enough appointment time they experience long periods of indirect waiting time (time between scheduling the appointment and that appointment becoming available).The patient also wishes to be seen either immediately or within minutes of their arrival (whether they scheduled an appointment or not). The time that the patient waits from the scheduled start time of their appointment to the time that they actually receive service is called direct waiting time. The patients using this method waste much unnecessary waiting time standing in line at the registration window to ensure a successful registration with a certain physician.

VI. PHASES



Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: Getting strong knowledge of the project title and doing exploration on it, we will get our description. Preparation: After we will first study about it and do some research on the project for our better understanding and also getting a rough depiction about what would be our problem definition for this particular project.

Designing: Then we will build the design of the project and agreeing to that, will list down all the necessities needed for the creation for the prototype of our project.

Phase 2: Testing and Deployment.

Testing: After the model is ready we will first associate the hardware with the allotted code and then we will check if it supports the machine or not. If not we will resolve the issues relating to it and will crisscross again.

Deployment: After completion of integration and testing of project, real time testing and operation of the system will be done.

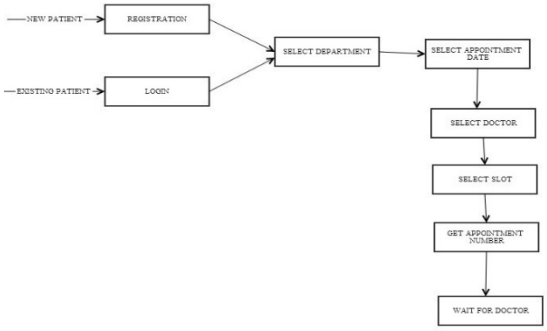
VII.METHODOLOGY

In online appointment scheduling system patients are given an appointment number. At the designated appointment time, patients arrive at the hospital and get the registration that is allotted to their appointment number. These patients need not queue at the registration window. In this way both direct and indirect waiting time can be minimized and hospital's valuable resources can be utilised efficiently.

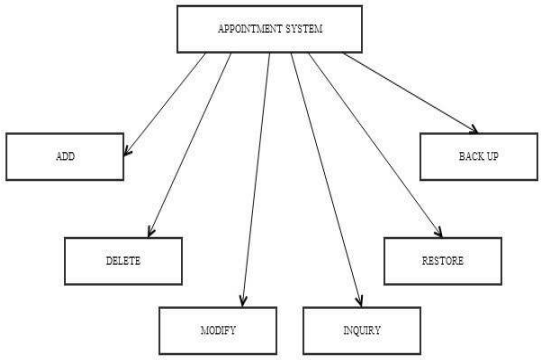
VIII. SOFTWARE ARCHITECTURE

- Schedule daily appointment list of patients
- View patients complete information
- Reschedule the appointment
- Assign time slot to each patient
- Cancel the appointment
- Check Doctors availability for patient's Appointment
- Send reminder SMS to patients

IX. BLOCK DIAGRAM



X.CIRCUIT DIAGRAM



Database Management consist of Appointment Details, Addition of data, deletion of data by the Authority, Modification of data

It also consist of inquiry section, as well as Backup

XI. OUTPUT

As we go on the appointment system, we get the option of registration as well as login option and then only we can go to the appointment section

XII. OUTCOMES

Creating the outline, works survey and achievability study for proposed solution which will act an input to the design phase. Suggesting Statement of work, scope description and scope border for planning the model from the problem definition to decide what needs to be done and what not to be done.

Phase 1 Implementation: Implementation of the online appointment scheduling system with the help commands running it in the Linux environment.

Phase 2 Testing: Testing was done with cases like testing mic and speakers , Testing by asking different queries to the assistant .The testing of the model was a success are mentioned in the above figures.

Phase 2 Deployments: Real time working was done when we asked questions and assistant replied the answers appropriately. Thus the model works efficiently.

XIII.RESEARCH GAPS

Every project has some gaps in it. No project is 100%efficient. There are some gaps in this project too. They are as follows:

Since the Wi-Fi was not working with the online appointment scheduling system ,due to this confusion was raised.

The system cannot control other devices connected on the network right now. It needs to be configured in that manner to do so.

XIV. SUMMARY

The user login and registration requests are processed by the portal server which located in 2- tier. The application server is a component that manages the complete end-to- end appointment tracking and scheduling services. The key functions of application server include: (1) multiple-practitioner scheduling, (2) centralized and consolidated patient appointment tracking, (3) available appointment searching, (4) appointments rescheduling, and (5) appointment confirmation and cancellation. Detailed information about each scheduled appointment slot, such as patient login and contact information, is also stored into the 2-tier database

XV.CONCLUSION

One of the biggest reasons that online appointment scheduling is getting popular day by day is that it helps the patient to make the appointment to their doctor, clinic or hospital in an easier way. It makes it through the computer, access a website or software and makes an appointment, than to go to the hospital, wait in a line for a number of hours, just to make an appointment with the doctor for the next week or next month. And through this, patients can also involve in the health decisions that they have to make. They can make an appointment to another doctor other than theirs, by nothing more than a click.

References

[1] Fatma Poni Mardiah, Mursyid Hasan Basri, “The Analysis of Appointment System to Reduce Outpatient Waiting Time at Indonesia’s Public Hospital”, Human Resource Management Research, 3(1) (2013), 27-33

[2] Cayirli, Tugba, Emre Veral,” Outpatient Scheduling in Healthcare”, Production and Operation Management, vol 12(4)

[3] Danielle Hayday,”Online Appointment Scheduling”, 2014

[4] Wenjun Cao et.al, “A web-based appointment system to reduce waiting for outpatients: A retrospective study”, BMC Health Services Research, vol 11, 2011

[5] Lucille Keir, Barbara A. Wise, Connie Krebs, Medical Assisting: Essentials of Administrative and Clinical Competencies, Cenage Learning pp.131-133

[6] Arthur Hylton III, Suresh Sankaranarayanan, “Application of Intelligent Agents in Hospital Appointment Scheduling System”, International Journal of Computer Theory and Engineering, Vol. 4(4) (2012)

[7] http://www.weberp4.com/industry- solutions/hospital-management/doctors- appointment-management.htm

[8] Xiuju Zhan, Xiufeng Liu, “Design and Implementation of Clinic Appointment Registration System”, Engineering, (5) (2013), pp.527-529

[9] Karen Davis, Stephen C. Schoenbaum, and Anne-Marie J. Audet,” A 2020 Vision of Patient-Centered Primary Care," Journal of General Internal Medicine, vol. 20, pp. 953-957, 2005.

[10] Xiaojun Zhang “Developing an Online Patient Appointment Scheduling System Based on Web Services Architecture”, APAMI Conference Proceedings, 2012

State Teacher Eligibility Test (STET) Software For Government Of Sikkim

Rohan Sharma, Urvi Sharma, Janhavi shetty, Pratishtha Singh, Smruti Singh, Rohan Vishwakarma
IT Department, Thakur College of Engineering and Technology, Mumbai
rohann20sharma@gmail.com, urvi321.us@gmail.com, janhavi9520@gmail.com, prati3331@gmail.com, singhsmruti2001@gmail.com, rohanvishwakarma@gmail.com

Abstract— **TET or Teacher Eligibility Test is an exam conducted at state levels to determine the eligibility of candidates for their appointment as teachers in central or state government schools. This exam is conducted offline (Pen and paper test) and it consumes a lot of time and labor. As a result we have come up with the idea of online portal of the same. The project entitled "ONLINE STATE TEACHERS ELIGIBILITY TEST" is aimed to conduct examinations for aspiring teachers in an efficient manner without time wasting for the checking of paper. Online examination system is a web-based examination system where examinations are given online either through the internet or intranet using computer system. . The main goal of this online examination system is to effectively evaluate the student thoroughly through a totally automated system that not only reduces the required time but also obtain fast and accurate results. In this automated system there is no need of paper and pen.**

Keywords – Online examination, Employment, Eligibility Criteria, Internet connectivity, Educational Institute.

I.INTRODUCTION

The Teacher's Eligibility Test is a test that is conducted for appointment of eligible teachers into Government Schools. This test is conducted for teachers from class 1 to 8 and is conducted as Paper 1 and Paper 2. Paper 1 for class I to V and Paper 2 for class VI to VIII. These days most states conduct their own TET. Which are conducted in offline or online mode. Today, Online Examination System is considered a fastest developing examination method because of its accuracy and speed. It is also needed less manpower to handle the examination. Almost all organizations today, are managing their exams by online examinations.

TET was introduced by Indian government in the year 2011 to improve the quality of teaching in government as well as private institutes. The teachers already teaching in the institutes were given time duration of 2 years to crack the examination for continuing their profession as a teacher.

The National Ministry Of Human Resource Development handed over the responsibility to Central Board Of Secondary Education (CBSE) to conduct Central Teacher Eligibility Test (CTET). Since, CBSE conducts this examination twice a year. On 8 December 2019, CBSE successfully conducted its 13th edition of CTET examination. This examination covers 110 cities across the country. Over 14 lakhs candidates participate in this examination.

In a country like India, where youths are in majority and are considered as the pillars on which our country's flag will stand to be on top of the world, quality education plays an important role. For our country's development in every aspect, it is necessary that the young generation should be provided with the best education this country has to offer. In such a scenario, teachers act as concrete to this foundation. A good teacher will give rise to an excellent student.

TET was made a necessary and important prerequisite for appointment of teachers all across the country to set a national benchmark for recruitment of teachers. It is also necessary to increase standards of teaching quality and improve performance of teachers, students and institutes.

We, by our project have given a solution to the Government of Sikkim to have their own Online State Teacher's Eligibility Test to ensure that the best and well qualified teachers get appointment as teachers in their state.

II. PROBLEM DEFINITION

Nowadays we have most states in India having their own TET But with the growing population and depleting resources the traditional pen and paper wouldn't be much helpful in the long run also, considering the ongoing digitalization every state is now opting for an online software to take up the task of successful and efficiently conducting the TET. One such government requiring an Online Software for TET is the government of Sikkim.

III. STATEWISE TET CONDUCT

Every year many states all across India conduct TET examination in offline as well as online mode for appointment of teachers. Statistics of this examination of few states are stated below.

A. Andhra Pradesh & Telangana (APTET): CBSE who is responsible to conduct CTET has given accreditation to Andhra University to conduct their state TET known as APTET once in a year for recruitment of teachers in their states. In the recent APTET held in 2019, over 4 lakh candidates from the state appeared for this examination.

B. Uttar Pradesh (UPTET): The Uttar Pradesh Basic Education Board (UPBEB) has the authority to conduct its Uttar Pradesh Teacher Eligibility Test (UPTET) annually. Mode of this examination is offline.

C. Rajasthan (RTET): The Rajasthan Board of Secondary Education has the responsibility to

conduct their state TET examination also known as Rajasthan Teacher Eligibility Test (RTET). Over 9 lakh candidates appeared for the RTET examination in the year 2019. The mode of conduct of this exam was offline.

I

D. Haryana (HTET): Board of School Education Haryana (BSEH) conducts Haryana's TET. The candidate who has qualified the exam has to reappear for the examination every five year. The candidate can give reexam end number of times. Over 2.5 lakhs candidates gave this examination in the year 2019 and only 21,993 candidates could crack it.

E. Punjab (PSTET): The Department Of School Education is responsible for Punjab State Teacher Eligibility Test (PSTET). In the year 2019 about 1.7 lakhs from Punjab and neighboring states appeared for the exam.

Sr. No.	Paper Title	Author	Year of Publication	Key finding	Research Gap
1	Web Based Online Examination System	Prof. Rahul Shahane	2017	This process is focused on removing unwanted vantages and is intellegible and user friendly.	In few Centres it was noted that only those who have registered already can give the exam and this process was done manually which is not appropriate.
2	An analysis on online examinations in college courses	Andrew P. Barkley	2001	Large cost saving of substitution of machines in labour. On spot evaluation of the exam and result display.	Unable to know the indepth theoretical knowledge of the teacher who is giving the exam.
3	Online Examination System.	Muna R. Hameed	2017	Obtain fast and accurate reluts in given period of time through a totally automated system using PHP and web servers.	The changes can be done in the question paper i.e. in an unauthorised way and the database can also be altered.
4	TET Online Examination	Shreya Sareen	2020	TET helps students getting proper knowledge from a educated professor. Professor also need not go to 10 schools or colleges and apply, rather they can give the exams and apply directly.	State TET is done for many states in India but not in Sikkim due to the network and runability issues of software at different places.
5	The Research and Design of Online Examination System	Zhang Yong-Sheng	2015	This system adopted B/S mode and used coding tool combined with MySQL database.	The area where there were network connectivity issues could not be able to give the exams which leads to unemployment of well-educated and hardworking people.

IV. LITERATURE SURVEY

1. Uploading the paper pattern and previous years' question papers on the website.
2. Issuing notices regarding the eligibility criteria to appear for the exam which requires the candidates to be aged between 18-35 and have a Bachelor's Degree in Education (or an equivalent course).
3. Declaring the examination dates.
4. Verifying exam applications which includes authenticating candidate's identification proofs and other important documents that make sure he/she fits the eligibility criteria.
5. Notifying the eligible candidates confirming their registration.
6. Conducting the online examination.
7. Providing the results immediately after the completion of the test.
8. Issuing online eligibility certificates to the passed candidates.
9. Allocating teachers to the educational institutes as per their requirements.

Following are the functions performed by the client:

1. Logging into the website.
2. Registering for the examination by uploading bonafide documents such as identification proofs, degree certificates, etc.
3. Appearing for the examination upon receiving the exam notification.
4. Obtaining result.
5. Receiving eligibility certificate.
6. Raising a grievance in case the candidate is not satisfied with the allocated institute.

V. SUBJECT DESIGN AND CURRICULUM

The website is based on client-server architecture. Here, the Ministry of Studies (Government of Sikkim) is the server and the teachers appearing for the eligibility test are the clients.

Following are the functions performed by the server:

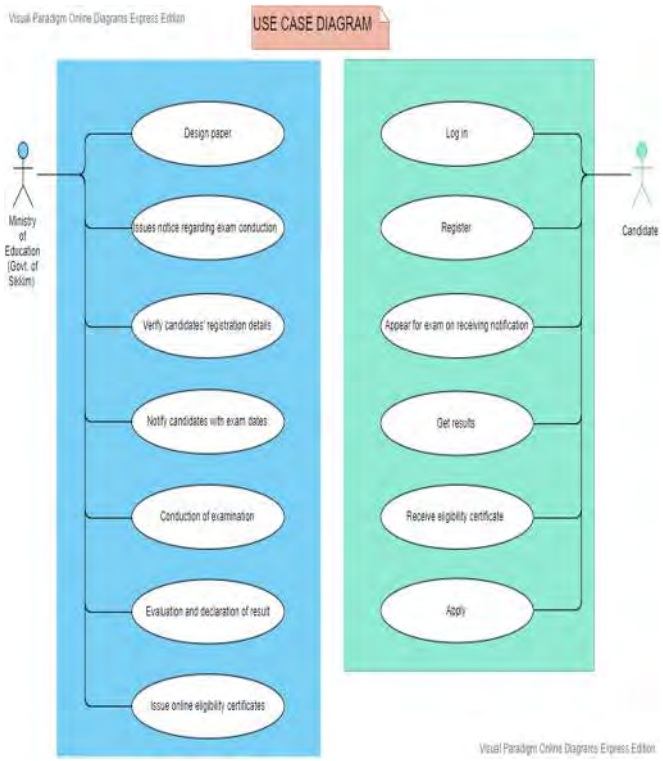


Figure : Use Case Diagram

VI. RESEARCH GAP

The online State Teacher Eligibility Test is conducted in many states across the country except Sikkim due to lack of network connectivity. The eligibility certificate of a particular state cannot be used in other state due to different eligibility criteria and requirements of subject knowledge. Since the mode of answer is objective in online tests, subjective knowledge is not tested which leaves us with a big question mark of whether that candidate is truly eligible or not. Due lack of knowledge of computers in many parts of the country lead, many candidates fail to understand the working of online examination and hence are left unemployed.

VII. EXPECTED OUTCOME

The main purpose of the system is to efficiently evaluate the candidate thoroughly through a fully automated system that not only saves a lot of time but also gives fast results and saves paper. It is a cost-effective and popular means of mass- evaluation system. The faculty prepares the tests and questions for each exam. The candidates can login through the client computers with their registration number given to them and can take the exam. The questions are shuffled in a random order so that possibilities for getting questions in the same order for the students who are beside is very less.

Objective of proposed system

I. Economic feasibility: When an exam is placed online, it results in significant cost savings. The cost of paper and distribution expenses are all reduced or eliminated. The elimination of paper costs alone is extraordinary. The copying and distribution of assignments to a large class are often unwieldy and inefficient. Administrators anxious to reduce expenditures are likely to strongly favor the transition from paper assignments examinations to using online assessment software.

II. Time Flexibility: Online examination systems make use of computers that helps in saving time. With the widespread availability of computers and the internet, there is a general acceptability and endorsement of this system. The lengthy formalities and processes involved in creating question papers, registering candidates for exams, answer sheet evaluation and declaration of results are completely eradicated with the online exam system. Each student is timed precisely and all results are generated instantaneously. In some cases, a candidate may even be able to get an assessment on failed questions.

III. Technical feasibility: A major highlight of using a web based exam software or an online examination system is that it gives a high level of transparency as opposed to the traditional method or remote method. It is almost impossible to compromise exam questions and evaluations because they cannot also be influenced. Most online exams generate their results instantly and it is often possible for the exam taker to get information on his results immediately.

IV. Eco-Friendly System: This is one of the major benefits of online testing. This kind of eco-friendly and forward-thinking approach to daily processes is essential in a world where students can graduate into an environmental crisis caused by climate change. From teachers, exam papers, electricity, etc, all of the cost can be drastically reduced by online tests.

VIII. SOCIETAL OUTREACH

The State TET will directly affect the employability of the teachers of Sikkim. As the teachers will have a proper government job after clearing the exam they would not have to roam here and there to 10 different schools and give all kinds of interviews, the teacher or professor will be all tested once by giving the test. This test is not only useful for the Faculties but also for the students who want quality education. Students in government schools and colleges need professors who are educated enough and can guide the students in the right path. Students will also willingly go to schools and colleges to gain education and stand on their

own. The 2 level examination procedure of STET i.e. Paper-1 and Paper-2 would clarify and distinguish between the faculties that who is eligible to teach higher standard students. The costing of papers and other materials would be cut off which was earlier used for the examinations. The technical growth of the state will also increase. Sikkim would turn more technical and would be able to compete different states at different levels. Since the teachers teaching would be well educated, this will attract the students from other states to study in Sikkim which will increase the economy and overall development of the State.

IX. RESULT

- The result affirms that our software helps overcome the drawbacks of the previous sites that conduct the online examination.
- In online examination process, security is one of the critical challenges. Proper authentication, authorization process is followed to ensure that the right candidate is appearing for the exam in a secure environment.
- In case of electricity cut, this software provides a data backup so candidate can continue the test from the same place as before.
- With improvising the software, we also contribute in saving the nature by building an eco-friendly system where there is no use or wastage of papers.
- Instant generation of results reduce the manual working of evaluation of test and saves the time for declaration of final ranks and further process.

X. CONCLUSION

The website provides facility to conduct online Teacher Eligibility Test for the state of Sikkim. It saves a lot of time, money and effort as paper cost is cut along with no labour charges to physically transport the question papers and collect the answers after the successful completion of the exam. The issue of electricity cut during the online examination may rise up but that is resolved using our software. The result gets displayed after the test is done which is generated by the server and this feature helps the examinee to not be burdened by stress of waiting for the result. The authority also has the privilege to modify and delete questions from the paper in case of any mistakes which ensures correction reaching on time. To sum up in short, the software developed, is a total win-win for both parties, the one giving the exam and the one taking it.

References

https://files.eric.ed.gov/fulltext/EJ1127649.pdf

https://en.wikipedia.org/wiki/Teacher_Eligibility_Test

http://www.academia.edu/Documents/in/Online_Examination

https://link.springer.com/article/10.1007/s10639-018-9797-0

https://www.shiksha.com/sarkari-exams/teaching/articles/what-is-tet-list-of-popular-teacher-eligibility-tests-including-ctet-uptet-otet-more-blogId-24515

https://link.springer.com/chapter/10.1007/978-3-642-33030-8_58

https://ieeexplore.ieee.org/document/7429241

https://pdfs.semanticscholar.org/3805/2b8204c61f7791286c0b085a26ac7ba3a2a3.pdf

https://pdfs.semanticscholar.org/3805/2b8204c61f7791286c0b085a26ac7ba3a2a3.pdf

https://www.researchgate.net/publication/317306939_Online_Examination_System

file:///C:/Users/Janhavi/Downloads/sp01ba02.pdf

https://www.collegedekho.com/articles/list-of-tet-exams-in-india/

E-Marketplace for Tribals

Vivek Shah, Laxmi Yadav, Rahul Singh, Anita Uniyal, Ashutosh Sanghvi, Riddhi Umap
Thakur College of Engineering and Technology
Mumbai, India
vivekmihirshah@gmail.com, laxmi31520@gmail.com, singhrahulkumar564@gmail.com
Anitauniyal2001@gmail.com, ashutoshsanghvi7@gmail.com, riddhi.umap17@gmail.com

Abstract:
As the trend in the urban areas consumers in the rural areas are also expected to have a hold on online purchase over time. Due to great advancement in technologies already the tribal people are getting access over better telecommunication service and witnessing increasing use of smartphones and computers. Taking advantage of these online portals we have come up with an application for the tribal people for marketing wherein people can do marketing over internet application or website. As 70% of India's population resides in rural area so this application can be useful. This shall reduce the human efforts and load which can serve the task of transporting the goods to the tribal by just relaxing at home. We humans have always looked for new inventions that meet our desire and comfort so this e-marketing for tribal can be beneficial.

I.Introduction:

Most of the forests and mountain regions are occupied by the tribes. Tribal people pay an important role in tribal areas of India by resorting to the sales of their own produce procured free of cost from the forests and mountain and making purchases of their other requirements.

Thus, they help distribute the forest produce to plain areas direct and through Government and other agencies, now they can also sell their
products online through E-commerce websites. The people are accustomed to live in small huts and houses.

Generally, they are below middle. Their cost of living is quite less when compared to that of the plains people. Low literacy rate is common in the tribal communities. The tribal people retain some of their sale amounts for investment required to enhance their business activity.

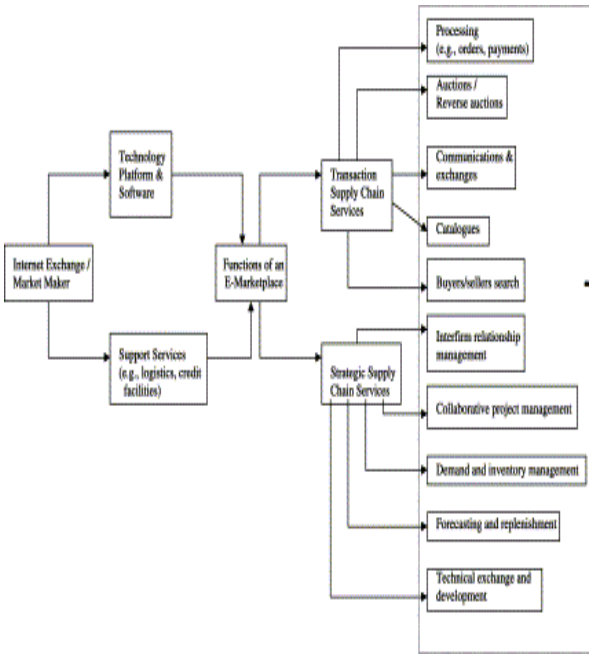
In addition to such savings, they get loans and advances from the Banks, G.C.C., Tribal welfare Department of the State, TRICOR, International Fund for Agriculture Development (IFAD), money lenders, middlemen etc. Financing business operations through these media is a very important aspect of development of business by the tribal people. The minor forest produce is procured and sold by the local tribal people in te weekly markets, along with that we can offer an online portal to sell those things using E-market or E-commerce website/applications

II.Problem Definition

The major wrong advantage taken by higher authorities is due to social backwardness, poverty and lack of digital literacy of tribal people. By which they are wrongly exploited and their trust on such schemes and great digital reforms are hindered. Due to which they themselves are made
to think that they are inferior from other sections of society and they are isolated from this great platform of digital development and transportation. Corruption is also biggest factor due to which awesome facilities and technologies are reaching to remotest of parts of India. Which ultimately doesn't serve the purpose of digital platform background
for every section of society. And this digital
isolation doesn't keep them attached and updated with happenings of world. Tribal regions are very well known for their sculpture, paintings, textile arts and pottery which indeed fascinate different sections of world. But because of literacy and lack of awareness of tribal people and wrong exploitation done by higher authorities the poor sections are made to sell their fine art products cheaply and they are actually are not paid the right price. Tribal areas are still thought of areas of wrong due advantage by world by which they are still thrown deep down in the black hole of under development and poverty.

III.Methodology

The customer first needs to register himself. The home page will contain the pictures of the recent searches of various products. Recent searches will appear with the help of keywords used. If any customer purchases article of more than Rs 500 discount of 10% will be provided.



Survey:

Sr no.	Title	Author
1	Journal Of Humanities And Social Science	Dr. Nirmala Alex1
2	E-Marketing-A literature Review from a Small Businesses perspective	Hatem El-Gohary
3	Tribal marketing: The tribalisation of society and its impact on the conduct of marketing	Bernard Cova and Véronique Cova

IV.Result & Discussion:

The following is the SWOT analysis of the tribal people adopting e-commerce in India:

Strength: Minor forest products are abundantly available in India, which can be used in productive and creative ways. Indian villages have a unique and diversified culture which produces a wide range of exclusive handicrafts. There is an increasing number of skilled artisans and an increasing knowledge about the demands of consumers, positive attitude of the sellers towards the government and non-government agencies and elimination of the middlemen due to e-commerce resulting in low distribution cost.

Weaknesses: Lack of adequate devices and infrastructure; lack of awareness about new designs, innovations and
technological upgradations; highly fragmented industries; limited capital and low investment capacity of artisans; insufficient market information on trends, opportunities and prices; limited access to credit and limited resources for production, distribution and marketing.

Opportunities: Global identity of the product and business and broader customer base due to e-commerce; rising demand from consumers in domestic and foreign markets; direct contact of customers and identity of self; middlemen being eliminated can make them fetch good price of the product in the market; government and NGOs' support and interest in preserving craft and marketing support.

Threats: The lack of authentic data on artisans, including their socio-economic status, livelihood conditions and family details is a major bottleneck which adversely affects planning and policy making for this sector; low infrastructural facilities, issues of logistics and network connectivity; decreasing supply of good quality raw materials; increasing customer sophistication and expectations; threats and exploitation by cyber intermediaries.

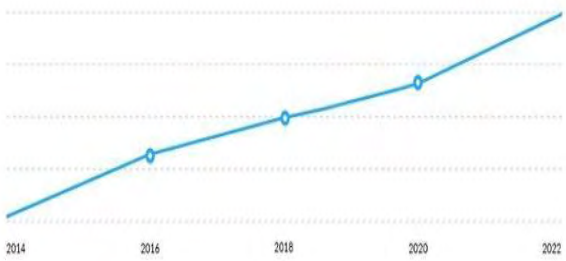
V.Conclusion

The Indian economy mostly depends on agriculture, as millions of people are directly orindirectly dependent on agriculture. Besides agriculture, rural arts and crafts give a gainful occupation to a large number of rural populations. They also serve as subsidiary occupation to agriculture and increase their rural income substantially. Major population of our country belongs to rural areas. Any strategy for national development must be associated with the development of rural areas. Every individual
artisan can develop his/her own website or with the help of cybermediaries, they can start e-commerce activities; NGOs' and other private institutions can help the tribal people to market their products online. Thus, e-commerce can provide tribal people an opportunity to reap the advantage of widened markets beyond the limits of geographical boundaries.

VI.Future Scope

Our biggest aim is to make a website to promote the tribal products so that we can overcome social backwardness, poverty and lack of digital literacy of tribal people. As we know e-marketing professionals are highly paid, so this market will increase job opportunities for tribal people as well. E-marketing has resulted in drawing the best possible revenue to the business. Due to high revenues, the discount on the products can be increased which will result in increase in sale of the tribal products. The Reviews on the product can be made mandatory so that product come in the search page and the product will stamp on the consumers mind. As we know the available e-market like Amazon does not sell ship product above 8 kg and does not sell product above 30 kg, we can incorporate the transportation facility to overcome this problem.

Global Tribal Survey for E-market



References

1) https://entrackr.com/2018/03/e- commerce-site-tribes-india/
2) https://nomada.biz/en/
3) http://www.csdindia.org/pdfs/Project- reports/TRIFED-Evaluation%20Study%20Report%2017-4-13-Final.pdf
4) http://www.iosrjournals.org/

PotSol

Nidhi Shetty, Avisha Jain, Aastha Shah, Hetansh Shah, Tanisha Rai, Sajjad Shaikh
Thakur College of Engineering & Technology,Mumbai, India
nidz251100@gmail.com, avisha9619@gmail.com, shahaastha8@gmail.com, hetansh403@gmail.com, tanisharainp20@gmail.com, starix456@gmail.com

Abstract— **Potholes on Indian roads are emerging as one of the biggest killers. The scenario is very dismal with more than 1,37,000 people dealing with injuries due to road accidents. Almost 30 deaths every day are caused due to potholes and speed which have claimed over 10,000 deaths for the last three consecutive years. Potholes Don't Just Cause Car Accidents, But Fatal Motorcycle Accidents Too. Potholes also cause major problems to the pregnant women and to other people too that are suffering from bone disorders. So we have come up with the solution called PotSol (Mobile Application for users and Website for admin). The aim is to easily manage the entire process of pothole management, hereby managing and dealing with potholes thus taking a step towards the betterment of the nation. Phases involving Centralized database generation, Image Processing, Pothole Sensing and Geo-tagging will be implemented and the target pothole area will be determined from the images that are uploaded by the users. Further the status associated with the tasks will help to manage potholes in an effective and efficient way thus preventing damage. Use of this proposed system will ease the entire pothole detection and its management thus minimizing the risk of human life.**

Keywords – Pothole, Geo-tagging, Image Processing.

I. INTRODUCTION

Potholes though seem inconsequential, may cause accidents resulting in loss of human life This may affect pregnant ladies as well as the people with cervical disorders. Potholes not only cause damage to the humans. But it also causes wear and tear to the vehicle. Which may lead to depreciation cost to the vehicle? P.M. Narendra Modi stated that "If Potholes aren't treated properly, they may lead to more deaths than cancer ". He has also started a development in the road sector again. He has grand plans to build more than 80,000 kilometers of roads by 2022. This paper, we present an automated system to efficiently manage the potholes in a ward by deploying Geo-tagging technique that overcomes the drawbacks associated with the existing survey-oriented systems. Every year, local authorities come under renewed pressure to maintain and improve the road infrastructure. Indeed, some would argue that public satisfaction with road maintenance is at an all-time low. However, with potholes on the rise, local authorities will have to consider how best to minimize their risk exposure and manage the budgets to ensure long term pothole and road maintenance. Proper risk management processes should therefore be put in place to mitigate both the reputational and financial damages associated with pothole management.

Figure 1.Literature Survey

II. PROPOSED SYSTEM

The proposed system consists of two modules application and a website. User Log's in and acquires the image of pothole. The uploaded image along with the Geo-co-ordinates is viewed by the admin. Admin views the pothole image and Geo-co-ordinates and assigns the tasks which will be visible to both the admin and the user. Also the status associated with the assigned task will be updated.

III. WORKING:

The Proposed system consists of an application (user) and a website (admin).

Images of the Potholes will be uploaded by the citizens through the application along with the automatically fetched location (through geo-tagging) and size (through image processing).

The uploaded image along with the Geo-co-ordinates will be viewed by the Admin (appointed by civic body) on the website.

Workers will be assigned tasks to resolve the potholes; also, the status associated with the assigned task will be updated which will be visible to both the user and the admin.

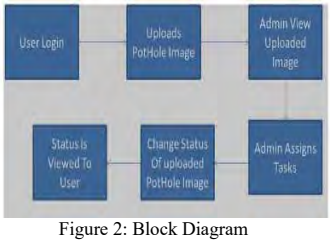


Figure 2: Block Diagram

Application:

- Register/Login: The user needs to register by providing the necessary information. If the user owns a vehicle he needs to provide its details too(for sensors). After successful validation with the help of two factor authentication, the user gets to see the different available options.
- Reporting a Pothole: To Report a pothole, the system provides two facilities:
  - While walking, if the user spots a pothole and wants to report it, the user needs to select the send picture option, click or choose the picture and click on upload; the appropriate location & approximate size are also attached with the help of Geo-tagging & Image Processing features.
  - While travelling, the sensors embedded in the tyres of the vehicles owned by the user will collect the information about the detected pothole and it will pass it to the application, which will then be uploaded.
- Check the Status of the reported Pothole: If the user wants to know current status of the grievance uploaded by him then he can check it by clicking on the Query status option. Here the user will also get to see the resolved pothole's Image provided by the officials.
- Citizen of the Month: The user reporting maximum number of potholes will be declared as "Citizen of the month" and may receive various goodies from the government.

Figure 3: Login page

Figure 4: Query Interface

Website:

- View Queries: Here the Admin(assigned by concerned government bodies) will view the received Queries through application along with the location on the map provided by geo-tagging and will assign workers to resolve the reported grievance.
- Providing Status: The time period will be automatically provided based on the size of the pothole and once the pothole is repaired ,the admin will provide complete status with the help of an image captured at the same location (so that the users cannot be manipulated). The status will be viewed by both user and the admin.

This System thus completes the loop, as the user gets a response to his reported grievance.

Figure 5: User queries

Figure 6: query details Interface

IV. Design Aspects of the System:

- Application (user)-

After launching the application, introduction about the system is displayed. For a new user, create account section and for an existing user, a login section is seen otherwise he/she directly sees the menu containing UPLOAD PICTURE , QUERY STATUS ,CHANGE PASSWORD and LOGOUT SECTION.

- Website (Admin)-

The website contains a home page where total counts about the number of potholes registered , resolved and pending can be seen along with the introduction of the system, it has a menu containing sections like view queries, view users, citizen of the month, logout and change password.

V. Advantages:

- Decreased number of potholes will lead us one step closer of becoming a developed nation.
- Number of Death rates will be decreased.
- Due to the feature of geo-tagging the user does not need to enter the address manually.

VI. Dependencies:

- Internet Connectivity
- Enabled GPS

VII. Conclusion:

Phases involving Centralized database generation, and Geo-tagging will be implemented and the target pothole area will be determined from the images. Further the status associated with the tasks will help to manage potholes in an effective and efficient way thus preventing damage. Use of this proposed system will ease the entire pothole detection and its management.

References

[1] https://www.telegraphindia.com/india/supreme-court-stunned-in-5-years-15000-died-in-road-accidents-caused-by-potholes/cid/1677940
[2] https://dictionary.cambridge.org/dictionary/english/pothole
[3] https://www.confused.com/car-insurance/true-depth-of-uk-pothole-problem
[4] https://www.thebetterindia.com/152644/pothole-free-roads-safety-india/

Commented [s1]:

Tracking Software with Hologram Technology to Secure Goods

Bhavanish Dhamnaskar. Harsh Sharma, Siddeshwari Patil, Dhanashree Shripatwar, Saakshi Pawar, Suyash Sawant
Department of Information Technology, Thakur College of Engineering and Technology, Mumbai
brdhamnaskar@gmail.com, (sharma.harshmukesh271@gmail.com, patilsiddi123@gmail.com, dhanashreeshripatwar@gmail.com, saakshipawar02@gmail.com, suyashsawant11@gmail.com

Abstract— **This paper presents a possible system for online shipment and delivery tracking using RFID technology and robust hologram grid system to ensure the security of the goods placed inside the container and protect it against theft. The software aims to track the location of the delivery at every instant of time, i.e. from the time of shipment to delivery, and using a smart hologram technology increase the accuracy in shipment of the packages as illegal attempt to unload any package would create an alert. So both the goods and their carrier vehicle are being monitored simultaneously. The database of the system is quite robust and provides useful information about the shipment by simply clicking buttons on the application which is available on both mobile and PC. Actual implementation of the software has not yet been tested but the results of the research show us that the system is promising for online goods tacking.**

Keywords: Hologram grid system, RFID technology, delivery tracking

I. INTRODUCTION

The need of tracking softwares is of a greater importance for many manufacturing firms nowadays. It is essential for managing logistics network efficiently and ensures high customer satisfaction. Delays in the shipment delivery of even a few days can cause a great loss to these high manufacturing firms. Coordination problems between shipment dealers and third party receivers causes imbalance in the production department as well as the sales department resulting in poor customer satisfaction. It also breaks the trust in relationships between firms gradually. Hence it is truly important for the third party receivers to know exactly till where their shipment has reached. Some companies have created great softwares used in tracking systems, but some limitations such as non functionality on iOS devices, lack of information displayed in mobile phones, networks issues causing failure in real time display of location. Also, due to minimum security arrangements for the cargo kept on the ports or in trailers or trucks, the risk of theft increases. Cargo theft may cause loss of about billions to the manufacturing firm. Till date no technological advancement has been made to overcome the problem of cargo theft. Without proper and suitable tracking and tracing system, efficient co- ordination of logistic flow would be difficult to acquire. Seeing all the above problems on hand, we have come up with an innovative solution to the tracking systems of the shipment of the goods and along with that ensuring the safety of these goods and ensuring the proper delivery of the goods. Our software aims to solve most of the problems faced during shipment of products by companies and also create a secure application for online order placing and buying or selling. More about the software is mentioned under the "Methodology" section of the Paper.

II. TECHNOLOGICAL REQUIREMENTS

Hardware: Trackimo (Tracking Device), Rrimin 3D Holographic Projector, Fingerprint Sensor

Software: Python, CSS, HTM, PHP, MongoDB, Netbeans IDE, Android Studio

III. LITERATURE SURVEY

Before developing the software design, we searched a list of firms who were making tracking systems. To understand what we should and what we shouldn't include in our software system, we researched about their tacking systems. We noted down the most common features amongst all the softwares of the firms. For any product, the most important thing is customer review, and hence we took the reviews of the customer and important part of the research. From there we got to know all the drawbacks the various softwares were facing. Some of these softwares are listed below:

1. Teletrac Navman Director

The satellite is used for the vehicle positioning by using satellite GPS. The information is transmitted between the vehicle and the telecommunications company (the security system) via a wireless cellular or device connection. The access to the information about your goods that are been tracked is easy.

Drawbacks:

- Satellite does not work in bad weather. Trackers can be disabled easily by easily disabling antenna.
- There is a lot of backend work that remains a big question mark to the end user.
- The system tablets malfunction on a regular basis.

2. ClearPath GPS

They are a no-contract company. The customers pay on a subscription basis for their GPS tracking system that can track vehicle location down to 30 second time-intervals.

IV. METHODOLOGY

V. PROS AND CONS

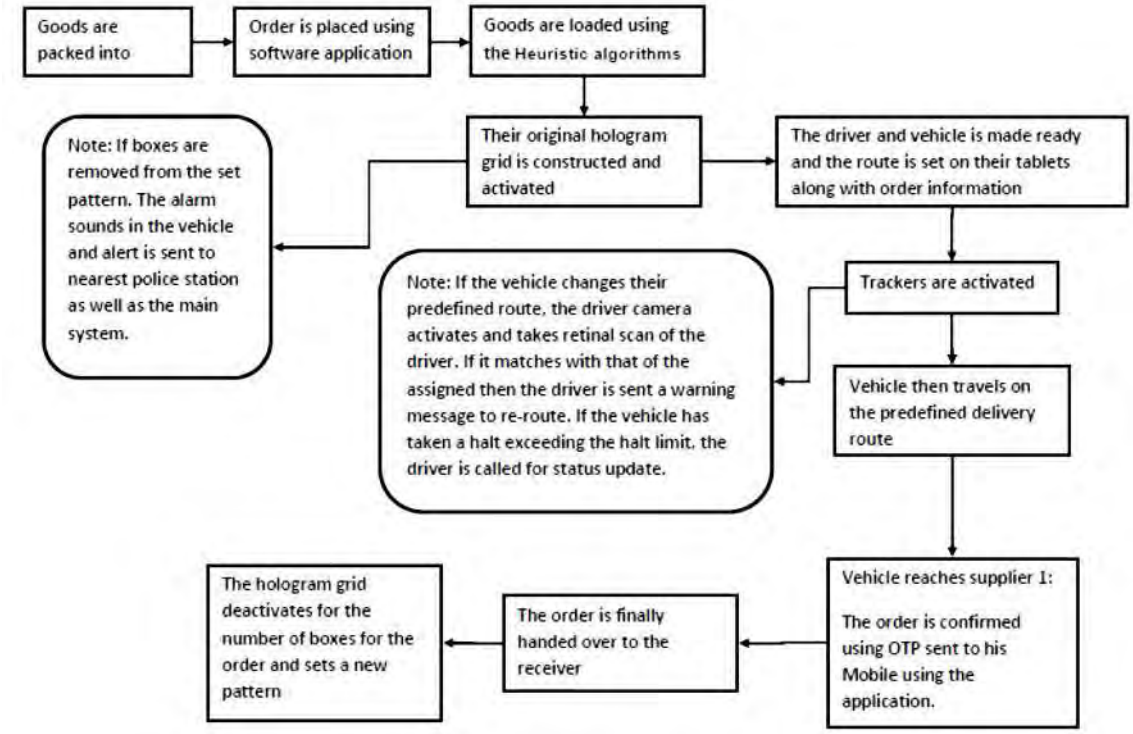
The benefit of having this software is that it not only tracks the goods along their delivery path but also omits the chances of incorrect cargo being unloaded. This will help in reducing the delays in shipment caused by incorrect unloading of the cargo. It also allows the customer to directly contact with the driver of the vehicle for any problems face. This software still remains helpless in avoiding damage to the cargo as the hologram just sets the alarm in case of movement but does not stop the cargo from moving. Also the hologram grid gets disturbed by the motion of the vehicle, causing false alarms at some times.

VI. FUTURE SCOPE

The future of this technology includes use of laser shields to protect the cargo against damage as well as theft. The limitations of the hologram system was that even after the alarm would be set on, the cargo could still be picked in case of robbery. Also the grid would be insensitive to the jerks of the moving vehicle which could reduce the chances of false alarms. Also offline mode for tracking the shipment would be introduced in the case of network issues. The tracking strength of satellite in case of densely clouded sky is still being worked on. This design promises a revolutionary change in the future of tracking technology for goods.

VII. CONCLUSION

With the help of such tracking software the



References

Development of an online platform for mentoring startups

Prince Pandey, Nandini Yadav, Aniket Singh, Pratyaksha Ambast, Vishal Singhania, Sudhanshu Rai
TCET,Mumbai
pandeyprince128@gmail.com, nandini976804@gmail.com, Aniket180101@gmail.com, pratyakshaambast2000@gmail.com, vishalsinghania52014@gmail.com, amanrai7757822267@gmail.com

Abstract- **Startup-a company founded by one or more entrepreneurs in an attempt to capitalize on developing a product or service which is unique and bring it into the market. For a startup to enter the major leagues, the innovation and uniqueness in their idea is a must. But the cases where a startup has an excellent idea but fails to make justice to it have been witnessed. Most of these companies either do not have a proper guidance or succumb to the external competitive pressures. The article focuses on providing the startups a mentor who can help them to mould their plan of action in such a way that it meets the industrial demand. The startups provide a short insight of their idea on the platform which the mentors and the potential investors can view and accordingly choose to act. Also the incorporation of Market Trends aid in the evolution of the company. We firmly believe the proposed application will greatly help in the startups to bring about a revolution in the existing industry,**

Keywords- Startup, Mentor, Investors, Market Trends

I. INTRODUCTION

The era of entrepreneurship has started and many students and businessmen are investing their time and money in developing many solutions to market problems which will loosen up the tensed situation in the market [1],[6]. The entrepreneurs need a guide/mentor to teach them the right way to approach to the system/business development and make it successful in the market. The past platforms haven't given all the facilities that a start-up should get and there was no existing platform for mentoring the start-ups online and providing them with a good mentor, in India. The proposed system solves this problem and has come up with a recommendation system which helps students to reach a mentor who will lead them with a purpose and help them with their technical, analytical and business knowledge along with their contacts in the industries to get funds or a sponsor. The stakeholders are the people who'll get benefited from this and time and money of the both parties will be saved by using this system.

II. LITERATURE SURVEY

Startup's are the great way to show out your individuality and create job opportunities for millennials, with the growing population of world with high percentage of technical background, mobile and internet domination have manifested in creating startups. Even the government has initiated for employment , entrepreneurship with 'Make in India' schemes for boosting development in youngsters. Developing sustainability is important in both internal and
external factors with disciplined planned route for the growth. To embark in the journey market trends , revenues, funding,venture are predefined factors to be taken into consideration. Infrastructure , team work, time investment, tenacity for work are the quality elements for entrepreneurial success. But for the development of an individual mentor plays an important role for business development and growth. Depending on the pesronality - traits and decision-making style mentors are needed for proper guide and support. Initiation,cultivation and definition
are much needed guide for startup to workout properly.
As far the reports more than 800 technology startup are being set up. Looking at the persistent growth rate of startups , incubators are much in demand, they act as cataylst
tool for development.
Break-up of Indian Startup BusinessesTechnology

1. E-Commerce - 33%
2. Engineering- 17%B2B - 24%
3. Construction-13%
4. Internet - 12%
5. Agri- products- 11%
6. Mobile apps - 10%
7. Textile - 8%

Challenges of startups to be meet by mentors [2],[4],[5]:

1. SaaS - 8% Printing & packaging – 8% Other – 13% Transport & logistics- 6%Outsourcing & support - 5%Others-32%
2. Challenges of startups to be meet by mentors:
3. Financial resources: Provided by Business developers
4. Revenue generation : Poor revenue generation can lead to failure of startup and due to this focus is deviated towards challeneges , funding, reduced revenues rather than fundamentals of business. capital growth is necessary along with sustainability.
5. Team members: Hiring team mates with talent for business and required soft skills is again an challenging thing. Founders of the startups need trusted members with potential skill set. Each member being specialized in their field of operations.According to the survey , 23 percent startups failed due to lack of member competency to handle the work efficiently.
6. Awareness in marketing : Marketing is ruling the leading world of IT , if market awareness is not created there are much chances of failure at first stage. Building community and firm is necessary for the establishment of the product.
7. Lack of mentorship: Most of the great startups have failed due to lack of proper guidance which is our main objective to implement the system. As an apprentice has no industrial knowledge, business and market trends for launching if product but have a great way for managemnet and implementation of

the product , it is necessary for them to have a mentor.

Already existing system [1]:

1. StartupIndia
2. MUDRA Yojna
3. SETU(Self-Employment and Talent Utilization)Fund
4. E-Biz portal

List of Current Startups and Area of Operations

Area of Operation	Startup Firm Name
Online food delivery	FRESHMENU,SWIGGY
Online fish, meat delivery	FRESHTOHOME
Big data analytics for trade	PEELWORk
Online pharmacy	MYRA
Platform to get local businesses online	NOWFLOATS

III. Research Gap

Through the studies we came to a conclusion that there's no such online platform being available for mentoring of startups as well as domain based filtering of mentors. Thus we propose an online platform for the same and let the upcoming entrepreneur to present their ideas and help them to get the mentors as well as financial assist through business incubators.

IV. Finance

The mentoring process can be a real stressful if the related topic isn't interesting to the client. That requires complete attention of the mentor and the entrepreneur to ensure that the project is being funded and implemented in the right way. Hence they require a perfect balance of time and money that is being invested in the project [9].

Now, the question arises that how will the idea get implemented with proper funding and who will sponsor/fund the project? The Businessman portal on the application will do the exact thing to satisfy both the parties simultaneously [12].

Fig.4.1 Finance

V. Methodology

The proposed application will have the following services/portals in it.

1. Startup – This portal will be focused on a startup's registration and will have the details like the domain, name, address and a mini article which will include the startup's basic idea. Depending upon a startup's mini article it will get the recommendations of mentors which are involved in the same field of interest.
2. Mentor – This portal maintains the details of the mentor. It requires the information regarding a mentor's name, field of expertise, address etc. The mentors will be initially approached as the heads of EDC cell of colleges. A start-up can enrol under multiple mentors and eventually select any one which they feel are the most suitable for them.
3. Market Trends- This will include market being divided as per the domains registered by the startups. With the help of experts there will be a data provided that will display the current business trends in the various market domains.

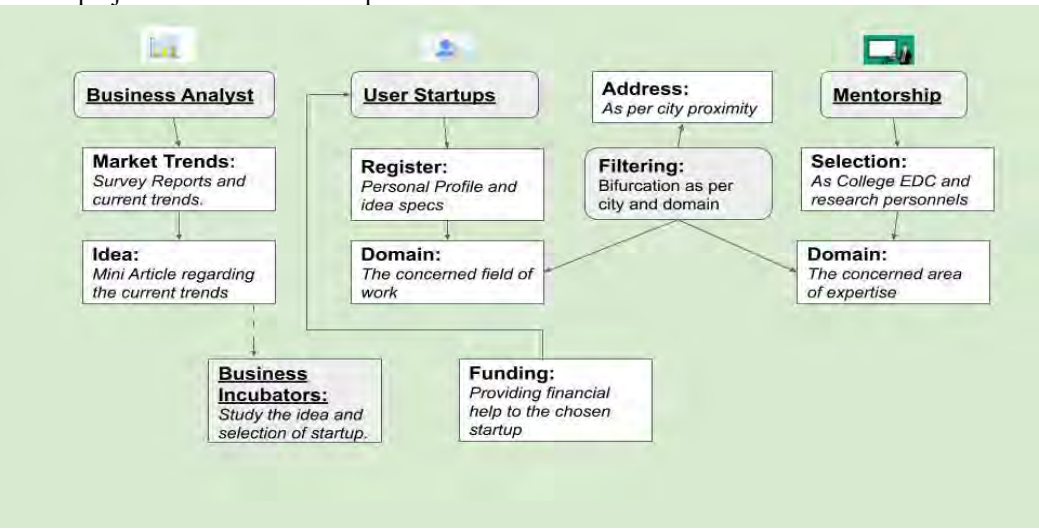


Fig.5.1 Methodology

This application can be used by the following user types

1. Startup personnel
2. Mentors/teacher guides
3. Business analysts

Table 6.1 Use case and scope

Scope

Startup	**Mentor**	**Analysts**
Registering themselves.	Understanding a startup's article and expressing their interest.	Study the market needs and instruct the startups to work in the pattern required by the industry.
Expressing a mini article regarding a startup's basic function.	Interacting with startups and help moulding them.	Interacting with the startups as well as the mentors.
Interacting with the mentors.	-	-
Understand the market trends in their respective domain.	-	-

VI. Results Achieved

Successfully Proposed an application which will act as a great aid for the existing as well as the newly found startups, and will incorporate all the key aspects right from the generation of an idea to seek a proper path for its evolution, a business acquirance opportunity to bring about a revolution in the field of interest and will be sustainable considering its future scope.

VII. Conclusion

The research proposes a conceptual framework for the evolving startups to seek mentorship in their respective domains. This propells a positive feedback as a technoprenuer may know the technology but a mentor ensures a startup to focus its attention on the society as well as the technology. The article also enables a startup to seek funding on an organised E-platform which can revolutionarize the current gruesome process and also the same time analyze the market trends in their field of interest.

References

[1]. Anand Paramjit, "Opportunities for Startups in India", Acreaty Management Consultant (P) Ltd ,The Entrepreneur,Feb,2016 www.theentrepreneur.com/article/270330 Accessed on 22.02.18

[2]. Chaudhary Varnana, "The Biggest Roadblocks Faced by Startups in India",2015 http://www.iamwire.com/2015/10/biggest-roadblocks-faced-startups-india/124312 Accessed on 24.02.18

[3]. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4]. Chokhani Rohit, Principal Founder, White Unicorn Ventures, "Challenges and opportunities for Indian start-ups; Key points to note", 2017. http://www.financialexpress.com/industry/challenges-and-opportunities-for-Indian-startups-key-points-to-note/524728/ Accessed on 20.02.18R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[5]. Derek Iwasiuk , "Key Challenges, Opportunities for Tech Startups in Emerging Markets," Moneycontrol.com,2016. http://ventureburn.com/2016/08/key-challenges-opportunities-tech-startups-emerging-markets/ Accessed on 22.02.18.

[6]. Grant Thornton Report- 'Startup India'- An Overview,20156. Griffith, Erin, "Why startups fail, according to their founders",2014. http://fortune.com/2014/09/25/why-startups-fail-according-to-their-founders/ Accessed on 20.02.18.

[7]. India filings Report, "Challenges faced by Startups in India",2016 https://www.indiafilings.com/learn/challenges-faced-startups-india/ Accessed on 19.02.18.

[8]. Mittal Ashish, " Indian Startups: Challenges and Opportunities",Economic Times, 24.11.14 https://economictimes.indiatimes.com/small-biz/startups/indian-startups-challenges-and-opportunities/articleshow/45272 Accessed on 21.02.18.

[9]. Nathani Komal"How this Start-up has Realized the Dream of Buying a Holiday Home for Many Indians", , Entrepreneur India, 2018 https://www.entrepreneur.com/article/308202 Accessed on 23.02.18.

[10]. Rediff.com, "20 good opportunities in India for entrepreneurs" http://business.rediff.com/slide-show/2010/may/04/slide-show-1-good-opportunities-for-entrepreneurs.htm#10 Accessed on 20.02.18.

[11]. Skok David , "5 Reasons Startup Fail ",Matrix Partners,2016. http://www.forentrepreneurs.com/why-startups-fail/ Accessed on 22.02.18.

[12]. Pandita Shivani, " 10 Financial problems faced by startups and their possible remedies", Knowstartup.com,2017. http://knowstartup.com/2017/02/10-financial-problem-faced-by-startups-and-their-possible-remedies/ Accessed on 23.02.18.

[13]. Aggarwal Aakanksha," Problems Faced by Startups in India and Solutions", Indianweb2, 2017 https://www.indianweb2.com/2017/03/10/problems-faced-startups-india-solutions/ Accessed on 2.02.18.

[14]. Telcom Regulatory Authority of India (TRAI)-A report ,2017.15. Times of India report, "17 startups to watch in 2017"

Centralized Attendance Monitoring System of Employees of RMDD

Abhishek Mishra, Aditya Shetkar, Ruchit Yadav, Hitarth Patel, Amar Singh

Abstract – **India, on its way of becoming a full-fledged developed country, still has some backlogs in some of its state's key functioning. The RMDD i.e. The Rural Development Department of Govt. of Sikkim has over 37196 workers. Maintenance of attendance of such a huge worker base goes a long way, but the pre-existing attendance systems have some flaws in it. For e.g. some workers tend to substitute anyone else if it's a card punching attendance record. Thus, our new proposal of a Centralized Attendance Monitoring system for those employees, would yield better results as well as help the monitoring officers to maintain, track and resolve any issues in the long run.**

Index Terms: RFID (Radio Frequency Identification), Buddy Punching, RMN (Registered Mobile Number), IRIS scanner, Authenticator code.

I. INTRODUCTION

Today, the attendance monitoring system is used globally. But the motive behind it is to be accurate about the employee's time of entry and exit. As we know, no project is completely flawless and full of pros. Similarly, some of the present and past attendance monitoring systems also have flaws in them. Our project: Centralized Attendance Monitoring System focuses on bypassing those flaws to make the regulation of such things smoother. The design and architecture of this project are such that there should be minimum efforts but maximum impact and output. Also, centralization should be an option available to every organization already using attendance monitoring systems. However, this project also emphasizes the fact that each and every entry marked is available to everyone and anytime within the authority.

II. PROBLEM DEFINITION

A. Review:

First of all, we must understand what **is RMDD of Govt. of Sikkim** to evaluate this problem. Ministry of rural development is a branch of the government of India, which is authorized with the work of accelerating the social & economic development of rural parts of India. It focuses on drinking water, health, housing, education, and roads.

The headquarters are located at Krishi Bhavan, New Delhi. The main objectives of this department ministry are sustainable growth of rural India through a multipronged strategy for eradicating poverty by increasing livelihood, opportunities, and developing infrastructure for growth. The department improves the quality of life in rural India and to correct the development imbalance.

There are currently **37196 employees** working for RMDD. Thus, the main problem sometimes arises regarding the attendance monitoring and updatation on the servers, which could sometimes lead to the compensation of a small number of salaries of people working in rural areas. The main issues in these cases are either of forgery/substitution (getting marked present while being absent) or some serious database issues in servers.

B. Actual In-Depth Problem:

In the 20th century, the need to accurately track employee hours has become an essential part of running a business. Laws were put in place to protect employees and ensure they were properly compensated. Businesses began using various methods of tracking employee time. While these methods allowed businesses to record employee time, they also possessed flaws. Here are the primary disadvantages of these traditional systems.

There is a Potential for Human Error:

When you rely on RFID or timesheets, there is always the potential for error. Employees may inaccurately enter their time into accounting software for payroll. These errors are common in the workplace when using manual time attendance solutions. The HR department would have to spend a large amount of time correcting these errors. The errors that go unchecked could result in incorrect pay for employees and other payroll issues.

Employees Can Commit Time Theft

Timecards/Timestamps and RFID do not prevent time theft or buddy punching. With manual time tracking, an employee could use another employee's RFID card to enter their time data in the system. This is known as buddy punching and is a common problem among large businesses, all around the globe.

Manual Time Tracking is Time Consuming

Handling RFID cards and timesheets takes a lot of time. The employees might need to stand in line to punch the RFID cards or timestamps in or out. The same amount of time is taken by the HR management to manage and maintain these records. This can include the replacement of RFID punch cards, correcting timesheet errors and mistakes, modifying the biometrics of an employee and other problems that arise when using a manual time tracking systems.

They are Inefficient and Outdated

The conclusion is that attendance systems like clock punch, timesheets, and individual data entry are inefficient

and outdated. They can have a large impact on any business or organization's performance. Time clock errors can cost the business time and money. These issues become bigger with large businesses, for example, any corporate giant. Not only for clock punch and timesheets but modern systems like RFID and biometrics can sometimes cause problems while punching, thus disturbing the relay time and inappropriate entries in the data servers.

Also, you can keep the disadvantages of using manual attendance systems solutions to a minimum when you only have a dozen employees. But, with a thousand or more employees, you could be losing thousands of rupees in payroll errors and time theft every month.

C. Background of Attendance Monitoring Systems:

A centralized attendance monitoring system (or CAMS) is a modern attendance management solution, nowadays used worldwide. Attendance management is the method or culture of managing attendance or presence in a work setting to minimize loss to the organisation due to employee downtime.

Attendance control has traditionally been approached using **timesheets, time clocks, and time tracking software such as RFID and Biometrics**, but attendance management goes beyond this to provide a working environment that maximizes and displays employee attendance accurately.

Attendance management takes place not only in all educational campuses but also in each and every single workplace, be it a corporate, business or factory. Nowadays, there are new types of attendance management systems like biometrics, RFID, GPS enabled systems.

These are critically important for any organization to function efficiently and smoothly, throughout its existence.

The first time clock was invented in the late 1800s with the purpose to record the time an employee entered and left the factory. This mechanical employee time clock would stamp day and time information on a thick paper card, hence the name 'time card'. This first time card gave the factory owner an actual record of the hours worked by each employee. This protected the business owner by making sure employees worked the number of hours they claimed, and protected employees by making it much more difficult for employers to cheat them out of their wages.

Since then began the era of attendance management, which has continued to date for decades and centuries.

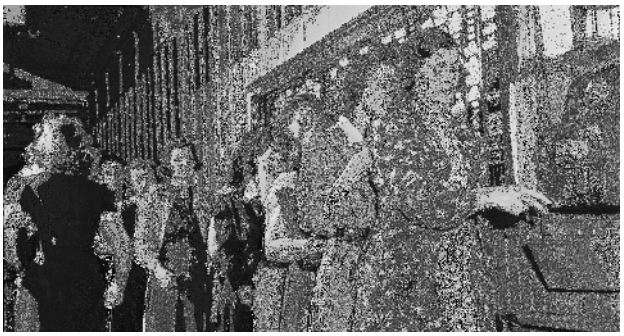


Figure 1: The First Employee Time Clock

D. Drawbacks of Previous Attendance Systems:

Drawbacks of some commonly used attendance monitoring systems are:

1. Time Clock:
a. Time Consuming
b. Time Theft
c. Human Error
d. Buddy Punching
2. Biometrics:
a. Expensive and High Maintenance
b. Complex Integration as compared to password solutions
c. Tend to be Inaccurate Sometimes.
d. Once comprised, it cannot be reset.
3. RFID:
a. Expensive
b. Costly RFID tags. Once lost, the employee has to repurchase the RFID tag which could be inconvenient
c. Prone to manipulation
d. There is a requirement to replace the microchip, radio transmitter, receiver and antenna, and battery which would be tiresome.

III. IMPORTANCE

Our Centralized attendance monitoring system mainly focuses on reducing data redundancy and data inconsistency. Since the project is for the RMDD: Govt of Sikkim, we would be integrating two or more types of a previously used attendance monitoring system to increase not only efficiency but also to increase the technical knowledge of the employees working in extremely rural areas to make them aware of the newest available technology. The following would be the main points on which the project would be mainly focused on:

- High efficiency and accurate time entries.
- Centralized, in order to make the head of the organization keep informed about the situations.
- Data backup on servers, in case of any emergency, arises.
- On-spot reset, delete or modifying options available with HR for convenience of employees.
- Enhanced Security.

IV. THE CAMS

A. Architecture:

The CAMS designed in this proposed architecture will have the following components:
1. Biometrics Scanner.
2. Iris Scanner.
3. RFID Scanner.
4. 2 servers hosted by central govt. of India
5. 1 server hosted by govt. of Sikkim.

B. The Working:

Since there were so many drawbacks in each and every individual attendance systems, we decided to collate them and make a comprehensive and complex, yet the most accurate and secure solution.
The employee whose data is to be entered would first place his finger on the biometric scanner. If it reads it successfully, then he may proceed to the RFID scanner. If the RFID reading process is successful, then his time-stamp is entered to the local database of the system in that office, which with no delay would be uploaded automatically to the state database server, which would be stored there for 14 days and then directed to the Central Govt's Main Server and Backup Servers.
In case, the employee is unable to complete their verification either by Biometrics or RFID, then he may proceed to the IRIS scanner. After they verify themselves with the IRIS scanner, they may enter the UID provided to them on their RFID and enter an authenticator code that would be delivered to them via SMS onto a computer screen which would be touch-sensitive. Thus, this proposed architecture is a 2-Step verification attendance system that is not only secure but would also have no redundancy in the entire process.
There is a necessity that the employee complete two verifications from any of the available options. The entire procedure takes less than 18 seconds to complete if there are no complications mid-authentication. These are the available options for the employees to validate their entry at the organization:

- Biometrics & RFID
- Biometrics & IRIS Scanner
- Biometrics & Authenticator
- RFID & IRIS Scanner
- RFID & Authenticator
- IRIS scanner & Authenticator.

The data entered on the systems & local database will be directly passed on to the servers of the Organization's current office which subsequently will be transferred to the state govt's servers, in order to protect the data from infringement and from getting it compromised. The Servers could either be hosted by govt or be rented from technological giants such as Amazon Web Services, Tata Technologies, Reliance, etc.

The employees would also be able to view their routine and performance with the help of an ERP system, which would be directly synced with the State Govt's records. Thus, any discrepancies noticed could be directly rectified with the HR management team within 14 days, after which the records would be uploaded to the Central govt's Main and backup servers. As soon as the state servers are updated with any discrepancy rectification, the Employee's ERP would also reflect the changes. Only the highest ministers of the state govt. would be able to view the status of the organization including the Authorities of the RMDD.

In case, any of the details of the employee needs to be updated in case of emergency or accidents, the HR of the local office could do so just by Resetting the particular field or entry in the device the employee needs to make changes in. For e.g, due to some accident, an employee loses their RFID. So, in that case, the HR can issue a new RFID along with a new authenticator code to verify at that very instant and then proceed with the issue. Let's take another example where an employee loses the eye which they registered as their verification for the IRIS scanner. The HR could simply delete that particular Employee ID' code from the database and propose the employee the option of either scanning the other eye for validation or not use that system at all.

Since the literacy rate of Sikkim is over 90%, implementing these systems in RMDD would not only increase the technology available in the area but also enhance their technological skills.

C. The Flowchart and Algorithm of working:

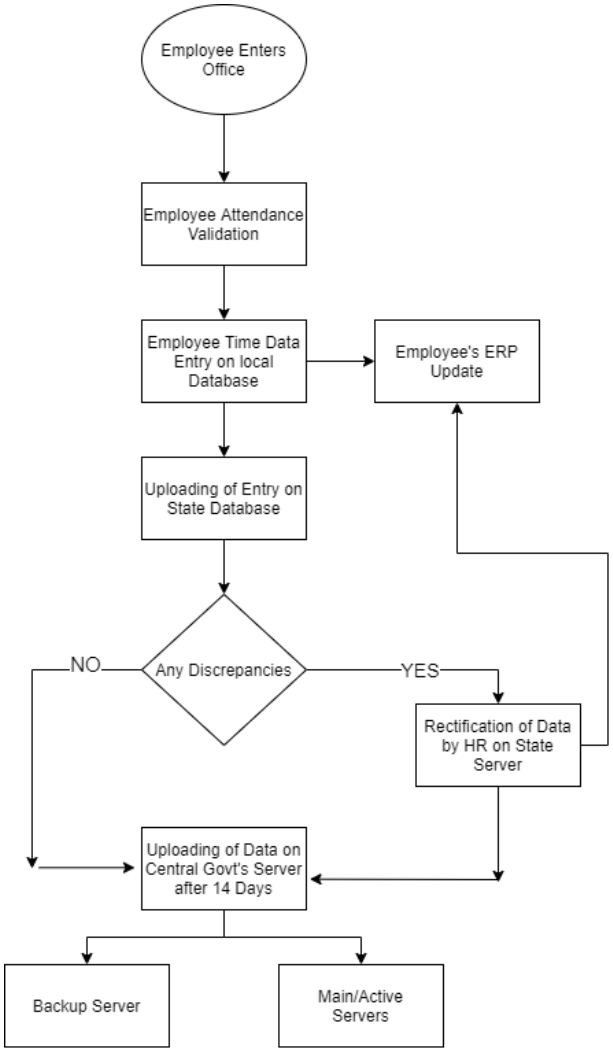


Figure 2: Main Flowchart

The above flowchart describes the general working of the Centralized attendance monitoring system. In the next flowchart, we have discussed the in-depth working of the validation process of the employee to enter their timestamp in the database.

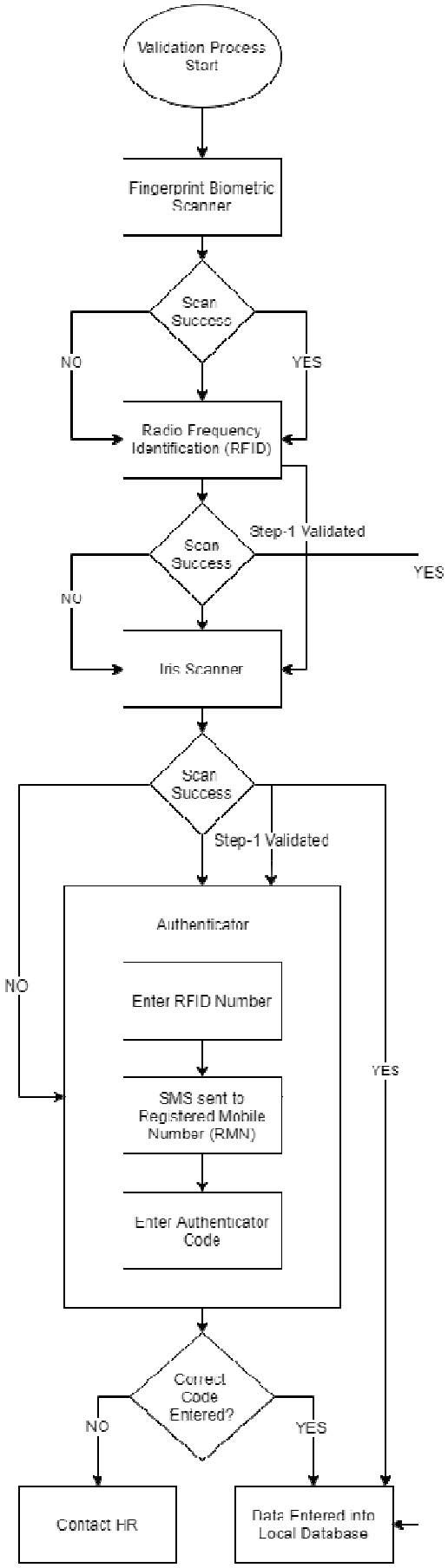


Figure 3: Employee Attendance Validation Process

V. HARDWARE & SOFTWARE

The Hardware proposed in the architecture are as follows:

- ZKTeco Biometric Fingerprint Scanner
- Matrix RFID attendance Monitor
- Cogent Radium Iris Scanner
- Touchscreen Inbuilt CPU Monitor.

The proposed Software to be associated with the hardware is as follows:

- MySQL Database
- Kronos Attendance Management Software
- Microsoft Windows 10
- Google Authenticator
- Microsoft Authenticator

The Total cost of this equipment would be around Rs.30,000/- INR considering the fact that the state and central govt's servers pre-exist.

VI. RESEARCH GAPS & DRAWBACKS:

As we said earlier, every project has some gaps in it and none of them are 100% efficient. Even our's has some drawbacks and these are the following:

- This project was designed not only for RMDD of Govt. of Sikkim but also for keeping in mind the corporate organizations. Thus, there could be some differences in the amount the organizations want to spend in a single attendance monitoring system.
- The hardware and software are expensive and so is their maintenance.
- The entire project revolves around the fact that both the state govt and the central govt own their own individual servers.
- The project would conclude efficiently, provided a good internet or broadband connection in the locality or the area of the RMDD offices.

VII. OUTCOMES

The outcomes of this project are expected to be positive as there might such a secure and appropriate solution for reducing redundancy, inconsistency and data loss. The operation of the project is simple, even though it's algorithm is a little complex. With this type of employee validation and verification, none of the employees could buddy punch the systems or manipulate the data to their convenience. Also, the central govt's data backup provides some additional safeguard measures for this project. The govt. of Sikkim had an annual budget of 7,051cr INR in FY 2018-2019. Thus even a small part contributed towards the IT team of RMDD could make this project possible.

VIII. SUMMARY:

A centralized attendance monitoring system is an attendance system that is linked to a central server and requires 2-step verification for the employee to mark themselves as a present. This is done in order to protect the RMDD govt. of Sikkim from employee's data tampering as well as manipulation. This 2-step verification is the best and the most secure type of attendance marking scheme. Also, if someone tries to delete some data about a month back, he couldn't do it, since the backup of all the attendance would be available in the central Govt's backup servers, which are installed in the first instance in case of some emergency.

IX. CONCLUSION

Thus, from a state budget of 7051CR INR, an efficient and accurate attendance monitoring system can be built. No organization in the entire world uses multiple linked attendance systems, due to which there arise problems of each individual system. Thus to cover those up, a combination of very those systems should be used along with some additional security measures.

References:

[1]. Bevan S., Hayday S. "Attendance Management: a Review of Good Practice" Archived 2007-11-09 at the Wayback Machine Report 353, Institute for Employment Studies, December 1998

[2]. Automatic Identification and Data Collection (AIDC) ArchivedMay 5, 2016, at the Wayback Machine

[3].Angell, I., Kietzmann, J. (2006). "RFID and the end of cash?"(PDF). *Communications of the ACM. 49 (12): 90–96.*

[4].Landt, Jerry (2001). "Shrouds of Time: The history of RFID"(PDF). *AIM, Inc. Archived from* the original (PDF) *on 2009-03-27*. Retrieved 2006-05-31.

[5]. FIPS-201, Personal Identity Verification (PIV) of Federal Employees and Contractors" (PDF). NIST. Archived from the original (PDF) on 2010-12-26. Retrieved 2019-01-10

[6]. How time and attendance technology can benefit your small business clients". Accounting Today. 2017-06-21. Retrieved 2017-07-05.

[7].Wayman, James; Jain, Anil.; Maltonie, Davide.; Maio, Dario (2005). An Introduction to Biometric Authentication Systems. Boston, MA: Springer London. pp. 1–20. ISBN 978-1-85233-596-0.

[8]. Reddy, P.V; Kumar, A; Rahman, S; Mundra, T.S (2008). "A New Antispoofing Approach for Biometric Devices". IEEE Transactions on Biomedical Circuits and Systems. 2 (4): 328–337.

[9].Funk, Wolfgang; Arnold, Michael; Busch, Christoph; Munde, Axel. "Evaluation of Image Compression Algorithms for Fingerprint and Face Recognition Systems". 2005 IEEE Information Assurance Workshop.

[10]. HRSID Iris Recognition | "rapid and reliable searches in both 1 to 1 (verification) and 1 to many (identification)"

[11]. "Biometric Identity Management System". UNHCR. Retrieved 2015-11-02.

Computerized Fee Collection System and Transfer Certificate Module

Vaibhav Kumar, Nidhi Gupta, Suman Gupta, Pratik Jain, Priyansh Jain, Jainam Jagani
TCET,MUMBAI
chikkalkarvaibhav@gmail.com, guptanidhiabc@gmail.com, sg1198959@gmail.com, pratikmjain17@gmail.com, priyanshjain625@gmail.com, jainamrj6@gmail.com

Abstract— **The Fee Collection Module is the one of the most automated fee calculation module on hand in the market. Apart from being automated it is, at the same time, flexible enough to accommodate the varying nature of fee payments that most of the educational institutions come across. Fee calculation is done on the basis of Category selected for the student.**

The module automatically calculates the pending fees, previous fee payment details, deductions and concessions if any applicable to the selected student. In addition to the above, the user is also given the option to allow payment of more than one month/fee-cycle either applicable.

User-defined fee structure and category can be maintained. Automated bill generation option enhances the system dependability. Fee Receipt can be accepted by the system with few mouse clicks. At any point of time, user can verify collections made and outstanding with single click.

For any school to register a child legally, it is mandatory that they receive a Transfer Certificate (TC) from the child's current school in which he is studying. Registration at any school without a TC is not possible. When a child is leaving a school or applying for a new one, he ought to have a TC for the future need and the school generates a TC with all the relevant details pertaining to the date of admission and date of leaving. The school also needs to maintain a record of the date of admission and fee due report of all the students before it could generate a TC. Hence Transfer Certificate Module will make one's task easier to generate transfer certificate without any delay.

Keywords–Fee, Transfer Certificate, Net Banking, School Administration

I. INTRODUCTION

In Fee Collection Module, you can effortlessly manage student Fee Collection
and maintain fees related data. You can view school wise total fee collected in a selected date range. This module displays total fee collected in particular date range by cash, card, cheque, Bank Transfer etc. You can also export the complete data to an excel file.

II. II. PROBLEM DEFINITION

Define different Fee groups for transportation Fees, Nursery to UKG Fees, Primary section fees (I to IV), Secondary section fees(V to X). Define penalty or late payment fees that get applied daily, weekly etc. Accept Fee in installments.

At a glance get to view the cash, card, cheque, bank transfer, Advance amount or disc. Obtain Fees in cash, credit card, NEFT, Cheque. Swiftly update enhancement in fees. Apply the different fees defined to the students as necessary.

III. III. BACKGROUND

When schools and colleges
are filled with stacks of loose papers and folders, you're overwhelmed with errors, late payments, slow pace of approvals, and forgotten invoices. Here is how to automate student fees so you can minimize careless mistakes and maximize your profitability.

IV. IV.OBJECTIVES

The specific objectives of this project include:

1. To provide a reliable and transparent system devoid of personal inclinations and interest
2. To provide a system that will relieve the staff and works of the academy the stress of searching files or data manually
3. The new system will bring an end to the issues of lost receipts or of any kind as the records and data will be stored in the new system's database which is safe and reliable for use and operation.
4. The new system gives a quick access to data and modification of records if necessary.
5. To save time (fast inputting and accessing of records)
6. To design a system that will be capable of generating the student receipt when payment is made and recorded with all the corresponding awaiting bills is incurred
7. To ensure accuracy in record keeping.
8. To ensure safety of the documents/records, because the system will be password aided to enhance security.

V. V. KEY FUNCTIONS

The main objective for developing this project is not only to achieve the targets and objectives set up by the computerized fee module but also to create the awareness of this programme among the various institutions.

1. Online Payment Method: The Fee collection module should furnish a provision to pay the school fee online. With digital transactions being

the current trend, every fee module should provide the above facility for online fee payment.

2. Integrated SMS Alert: The entity should be integrated with SMS feature i.e, timely alerts should be sent to parents about the due date and also an intimation or notification upon successful fee payment.
3. Receipt Generation: It should be able to generate an automated receipt after fee payment and also generate previous fee receipts. It should be compatible with both laser printers and dot matrix printer enabling you to use your
4. existing print technology rather than investing in new one.
5. Mode of Payment: The system should offer different modes of payment to parents like payment through debit card, credit card and net banking. It should allow the parents to pay the fee at their comfort and feasibility.

VI. VI. PHASES

Phase 1: Planning, Analysis, Designingand Implementation.

Analysis: We will collect the data about different fee collection transaction histories and Transfer certificate logs from data repositories of different schools and analyze them.

Planning: After we finish the analysis, we will use the statistics obtained, get a defined result and then build a plan based on the result.

Preparation: Based on those results, we will conclude which domains of educational institutions need more emphasis and will try to prepare the website on that real.

Designing: We will build the design of our project according to our analysis and accounting to that, we will list out all the necessities required for the creation of the prototype of our project.

Implementation: The design can now be implemented through web application frameworks like AngularJS and the required databases can be handled through software like MySQL.

Phase 2: Testing and Deployment.

Testing: After the prototype is ready, we will test the UI/UX of our website and will try to ensure that it is feasible with our targeted audience and creates the impact as expected.

Deployment: After completion of integration and testing of the project, the project will be deployed with real-time databases and real-time & operation of the system will be done.

VII. VIII.HARDWARE AND SOFTWARE REQUIREMENTS

Hardware: Any computer System or smartphone.

Software:

IDE Eclipse JAVA

1. Front End

a. HTML

b. CSS

c. JAVA

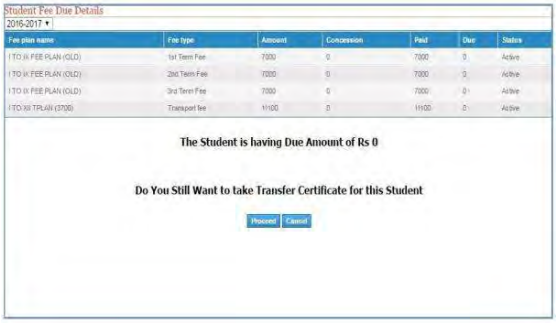
d. JAVAScript

2. Back End

a. MYSQL Database

b. PHP

VIII. IX.PROPOSED UI SCREENSHOT



TRANSFER CERTIFICATE

Affiliation No.:130105 School Code:06158

Book No.: Sl.No.: Admission No.:NIL24

1. Name of the Pupil: VEDIKA RAJPAL

2. Mother's Name:

3. Father's/Guardian's Name: PRUSHOTHAM S

4. Date of Birth(in Christian Era) according to Admission & Withdrawal Register (in Words & in Figures): Thirteen -November-Two Thousand 1(13.November

5. Nationality: Indian

6. Whether the candidate belongs to Scheduled Caste or Schedule Tribe or OBC: OBC

7. Date of first admission in School with class: 17 June 2014

8. Class in which the pupil last studied(in figures and in words): Class ISec F

9. School/Board Annual examination last taken with result:

10. Whether failed,if so once/twice in the same class:

11. Subjects studied:

12. Whether qualified for promotion to higher class (if so,to class(in fig & words): Yes

13. Month upto which the pupil has paid School dues:

14. Any fee concession availed,if so, the nature of such concession:

15. Total No. of working days in academic session:

16. Total No. of working days pupil present in the school:

17. Whether NCC Cadet / Boy Scout / Girl-Guide (details may be given):

18. Games played or extra-curricular activities in which the pupil usually took part(mention achievement level therein):

19. General Conduct:

20. Date of application for certificate: 01 Feb 2017

21. Date of issue of certificate: 01 Feb 2017

22. Reason for leaving the school:

23. Any other remarks:

Signature of Class Teacher | Checked by (with full name and designation) | Signature of Principal with date School Seal

The Fee Collection page is designed as follows which shows the fees collected and fees paid. The Transfer Certificate page shows the details already entered from before and the process is a single click application for submission.

IX. X.OUTPUT

The focus of our project lies on the smooth working of educational institutions by providing better solutions to avoid haphazard situations related to fee collection systems. Also provide a better plan to develop software for the generation of transfer certificates. Users of the fee collection software will have a platform to raise issues related to delay in fee submission, to ask for extensions,etc. directly to the authorities in-charge.

X. XI. OUTCOMES

Phase1 Implementation:
Implementation of a website with proper database and no failure.

Phase 2 Testing: Testing was done considering the feasibility of UI of the website among the users and the portability of the website on different devices

Phase 3 Deployment: Real-time working was done with teachers and parents from different schools and they were found to be comfortable with our website, they shared good user experience. Thus, the project works efficiently.

XI. XII. RESEARCH GAPS

Every idea was once a dream, so research gaps go hand in hand. The research gaps in this project are as follows:

- Due to poor connectivity this system may not work efficiently for rural areas. Websites may load at a slower pace and may pop errors a lot of times.
- There will be a need to provide training to parents as well as educational authorities and to all the people that will access the website so that it functions smoothly.
- All the people dealing with the website need to have a basic knowledge about how transaction is done safely to avoid blunders.

XII. SUMMARY

This website will act as a platform between the parents of the students and the school administration staff. The main purpose of this website is to make the fee collection easier and efficient while also making it easier to auto-fill in details while filling out transfer certificates for the students. It will also contain all the information of the school administration and general information such as school location, timings, etc. This website is purely made to decrease crowds in school during the manual fee collection timings and to make it easier to pay online.

XIII. CONCLUSION

We have successfully shown the implementation of our website. Technologies used to design this COMPUTERISED FEE COLLECTION SYSTEM AND TRANSFER CERTIFICATE MODULE are mentioned. Thus effective execution of the project is done. The concerned data will be reported to the concerned person on time and all should get the benefits of this computerized module.

References

[1]. https://www.conductexam.co m/fees-management-system
[2]. https://www.myclassboard.co m/fee-collections/
[3]. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ©IJRASET 2013: All Rights are Reserved 806 "Online Fee Payment System"

Software Tool to Aid Search and Rescue of Aircrafts

Shubham Jain, Aman Gupta, Hridyansh Gupta, Dev Jindani, Manu Khandelwal
Thakur College of Engineering and Technology
Mumbai, India
Shubham.Aryanjain@Gmail.Com, Avgupta3100@Gmail.Com, Hridyanshgupta@Gmail.Com, Devjindani68@Gmail.Com, Manukhandelwal20@Gmail.Com

Abstract— **Aircrafts are a boon for mankind, allowing us to travel long distances in short time, but they do come with certain risks like pilot error, mechanical failure, harsh weather, sabotage and other types of human errors. Thus, it is very important to create a software tool to aid in search and rescue of aircrafts. This paper deals with the features of software tool required to aid in search and rescue of aircrafts.**
A software can be created which inputs the flight route of the aircraft and stores details like fuel left, average speed of aircraft before it went missing, altitude.
An enhanced black box which communicates the inputs mentioned before to the software at fixed intervals is developed.
Thus, the software takes into account all the inputs and weather factor and runs various algorithms and provides with a probability circle of position of missing flight.
Hence knowing a close approximate of the missing aircraft will help in speedy search and rescue operation.

I. INTRODUCTION

"Aircraft search and rescue operation priority is at a all-time high "

This paper basically deals with the problems faced by the search and rescue team of airport authority of India. Firstly, the question is" What is a search and rescue (SAR) team?". Search and rescue team is basically a group of professionals who work on helping and rescuing the people on board a flight which is in trouble during crisis such as plane crashes, plane missing and highjacks and work on tracking down the missing aircraft.
SAR teams today have various software but they are in-efficient to provide the probable location of aircraft which increases the search time and thereby delaying rescue.
Designing an efficient software requires some thought process about the use and specific needs of SAR teams. So, there are some steps and rules to be followed which will be discussed later in this paper in designing a search and rescue software Designing an efficient software is very significant because it needs to meet the expectations of various aircraft agencies and governments.
The software created should have the potential to be equipped with future technologies and not become obsolete.

II. PROBLEM DEFINITION

The primary problem faced by the search and rescue team (SAR) of airport authority of India but also by most of the airport authorities is that they are unable to track down the location of the missing aircrafts within timespan required to execute their rescue operation. Due to this many of the aircrafts have gone missing, leaving no traces which indirectly affects the families of the people onboard and also the reputation of the country leaving everyone in distress. To resolve this problem, we can create a software that predicts accurately the location of the aircraft and during the time of distress it shows the last detected location of the aircraft and guides the SAR team to the most probable location of the aircraft.
Thus, creating a software which overcomes the limitations of previous software used by SAR teams is the desired solution.

III. BACKGROUND

Researchers have developed technologies that has reached new heights but still need work. Airport authority of India and other countries have their search and rescue teams which are doing their best with the resources they are provided with. They can achieve a higher success rate by working with better and improved technologies. The basic goal of every search and rescue team is to complete their task successfully and execute the rescue operation and save as many lives as possible, but they are not able to do it with a higher success rate. SAR teams also have to provide with answers as to why they were not able to perform more efficiently and the answers all point to one thing i.e. time-consuming search process.

IV. IMPORTANCE

It is very important for an airport authority of every nation to provide comfort and safety to every passenger on board. But during a time of crisis i.e. during a crash

or sabotage it becomes very difficult for them to regain the access to the missing aircraft or to locate it at the right time and rescue each and every passenger safely. Sometimes due to harsh weather conditions or due to some technical reasons the pilot or the particular aircraft loses its contact with the air traffic controller. It is very much important to develop a kind of software aiding tool that predicts the last detected location of the aircraft so the SAR teams can work on it immediately and help everyone in distress or evacuate the area if unfortunately, the aircraft has to crash land.

V. OBJECTIVES AND SCOPE

1.Previous software used by SAR teams do not efficiently aid in locating the aircraft hence the software proposed greatly reduces the time and manpower required for search operations.

2.Though there is a decrease in rate of the missing aircrafts, many of them are not found.

3.We only find the remaining parts of aircrafts which wash onto the shores.

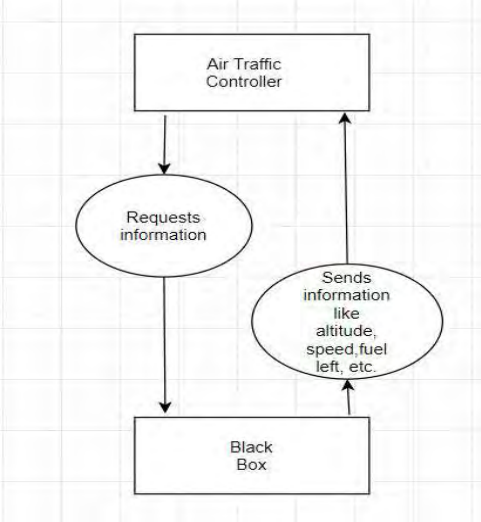
4.The proposed software greatly reduces time consumed to search the missing aircraft and start the rescue operation.

5.The software provides the probable location perimeter of aircraft when it goes missing taking into account the amount of fuel left, aircraft speed, wind direction, altitude and many factors.

6.The main objective is search and rescue and the most important factor while travelling by an aircraft is tracking which the software ensures.

VI. DESIGN OF SOFTWRE

A software basically consists of a front end and back end. The logic of the software is written in the backend and the frontend which is basically what a user sees and operates on. The user interacts with the frontend and then the logic is worked upon and the data is linked to the backend hence providing the required information. This software consists of a link which connects it to the chip in black box located in the aircraft, which provides the software various details like speed, altitude, overall health of aircraft, route of aircraft, turbulences faced and the conversation the pilots are having. The GPS chip and software combined in the aircraft provides the live and predicted locations of the aircraft. This can be linked by satellites to provide accurate and satellite images of the probability circle of plane crash. So when a crash occurs the application shows the last position of the aircraft runs algorithms and various simulations using the details provided by the black box and outputs the probability circle of plane crash and the SAR team can trace down the location and execute the operation thus saving much time as compared to earlier practice



The following points must be considered while designing the software –

[A] *Selection of Platform*

The usage of the software cannot be restricted to a single platform, this software should be made such that it can be operated on any system such as mobile phones as well as computers, tablets.

[B]*Selection of Software*

Java is primarily the most used language to create an application for various needs. However, other programming languages such as Python, PhP, Ruby are also widely used to create robust applications. For front-end JavaScript is preferred.

[C]Compatibility with the system

The software should give the appropriate details required of the aircraft using location trackers, GPS, etc. The system should be able to work in symphony with the back end with least possible lag.

VII. IMPLEMENTATION PROCESS

[A]*Creating the frontend*

All the essential packages required for creating the frontend are imported using the required programming languages, the black box is set into the aircrafts. On the frontend there will be an option to enter the flight to be tracked and by confirming the details all the required information is displayed to the authority of airport.

[B]*Creating a database*

All the necessary details are mentioned on the database in the backend which is linked to the application ,so when the required search is made all the details are tracked down, it will the calculate and plot the area where the search should be conducted from the last known location of the missing aircraft using the flight details and information provided by the black box.

VIII. SOFTWARE REQUIREMENT

There are various tools that can be used to create this software like the basic compatibility systems,the GPS trackers in blaackbox so that we get the live location continuously and this software can be implemented on the system of airport authority of India also the other airports.

IX. RESULT

Most people are scared to use airways as an option of transport mainly due to safety issues and due to some crash incidents reported in the past. Before it was very difficult for the search and rescue team on Indian airports like : Antonov An-32 twin engine turboprop transport aircraft of the Indian Air Force disappeared while flying over the Bay of Bengal... The search and rescue operation became India's largest search operation for a missing plane on the sea in history. This problem persists not only in India but also for an other instance in past: Malaysia Airlines Flight 370 disappeared on 8 March 2014, after departing from Kuala Lumpur for Beijing, with 227 passengers and 12 crew members on board. Malaysia's former Prime Minister, Najib Razak, stated that the aircraft's flight ended somewhere in the Indian Ocean, but no further explanation had been given. These aircrafts were not found at the right time or were never found due to lack in the resources and technology, through this application the exact search plot of the aircraft can be determined so that there is no confusion and rescue operations can be done on time and successfully and there are no further casualties.

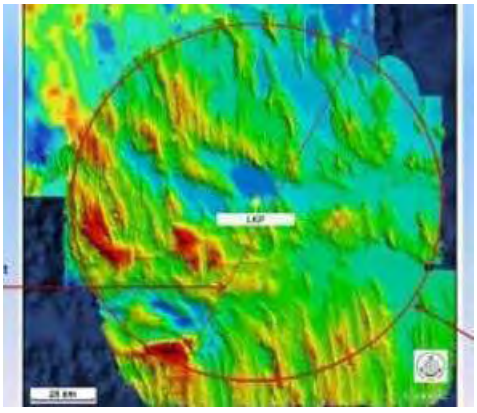


Fig: Probability circle of aircraft crash

X.CONCLUSION

We have successfully shown the importance of this software, that can be implemented by the airport authority of India ,that can act as an aiding tool for the search and rescue team to conduct the search when an aircraft goes missing by plotting a particular area where the search should be conducted so that the team reaches that area quickly without wasting the essential time in searching the whole area instead. This will help in the search and rescue operations very efficiently. Thus the importance, need and working of this software is successfully implemented and understood.

References

1)https://www.skybrary.aero

2)https://www.geeksforgeeks.org

3)https://en.m.wikipedia.org/wiki

4)https://www.icao.int

Alumni Tracking System

Anurag Mishra, Pranali Darekar, Ashutosh Mishra, Kripa Damania and Raj Jadhav

Abstract – **The information system for communication with alumni embodies one of many ways how university can keep tracking with its graduates. Except for communication between university and its graduates, the information system should allow communication between graduates themselves and their personal presentation in public. The system also should collect actual information about working experience of graduates, which can improve faculty credits and teaching process. The presented information system includes all these points and focuses on usability and comfortable user interface.**

The aim of this project is to build an Alumni Tracking System university online dashboard, it is intended to manage especially the Alumni of college with their respective fields.

The project mange the fresh as well as old graduate students with their respective information in actively participating in making registering, searching, managing the alumni information for sharing their expertise, network, jobs opportunities and resources. This participation has evolved into multiple dimensions; and is now eager to formally enter into the domain of mentorship to students through a regular program. Alumni registration management system (Alumni Dashboard) is a step towards this, creating multidimensional interactions between current and past students of College

I. INTRODUCTION

The greatest asset any Institution can have is the Alumni system. Alumni are the people who represent the Institution in the real world.

Alumni website is created for the students that have graduated from the Institution. This is a web based application that allows former students to take advantage of the benefits and services that a Institution offers after graduation. The alumni network is becoming important in the development of the institution because of their vast potential that benefits both the Institution and the students. There are many benefits for being an alumni member of a college or Institution, some of these benefits are: keeping a person inform on the events that are organized by the Institution, and when some important events will be holding in the Institution. Another benefit is that information concerning a former student can easily be received and other members of the alumni community can be located without much stress. The student and alumni can communicate each other.

II. LITERATURE REVIEW

This project is aimed at developing a Repository and each Engine for alumni of the college, which is of importance to a college. The Alumni Information Database is a web based application that can be accessed throughout the World. Anyone can access the search Engine to know about any Alumni of that college .

This system can be used as an application for the Alumni Information

This project is aimed at developing a Repository and each Engine for alumni of the college, which is of importance to a college. The Alumni Information Database is a web based application that can be accessed throughout the World. Anyone can access the search Engine to know about any Alumni of that college but can't able to add.

This system can be used as an application for the Alumni Information

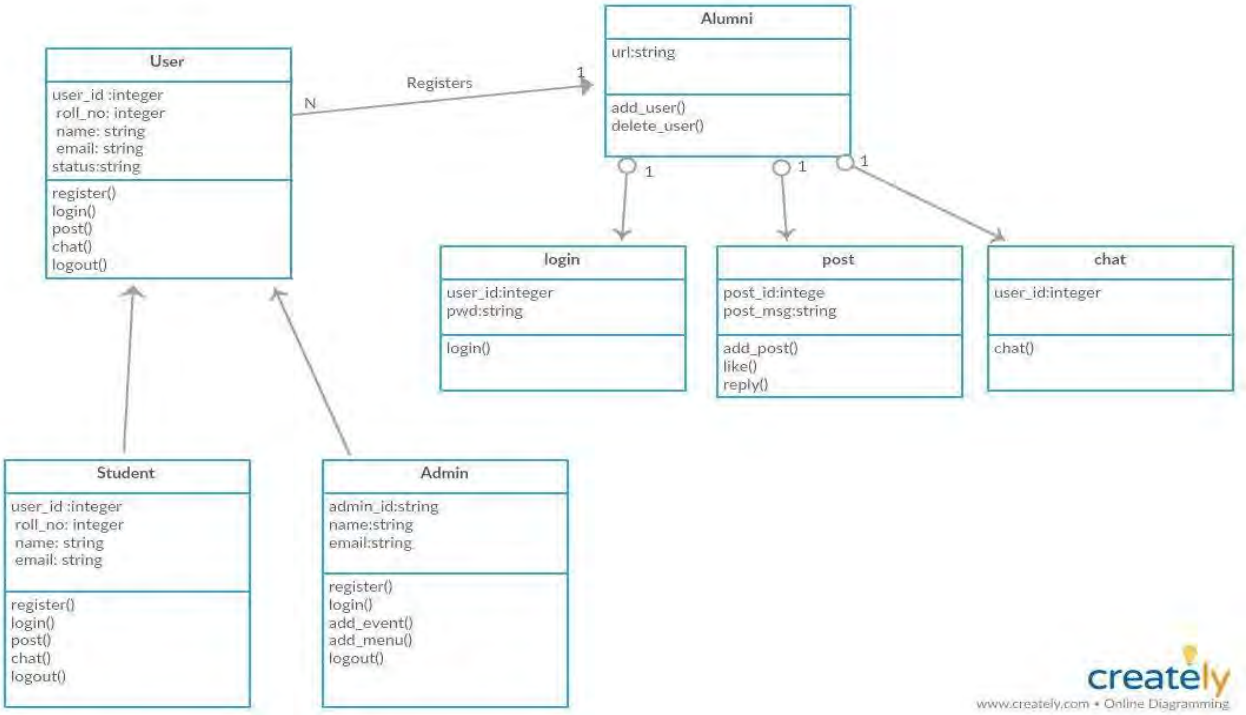
III. IMPLEMENTATION

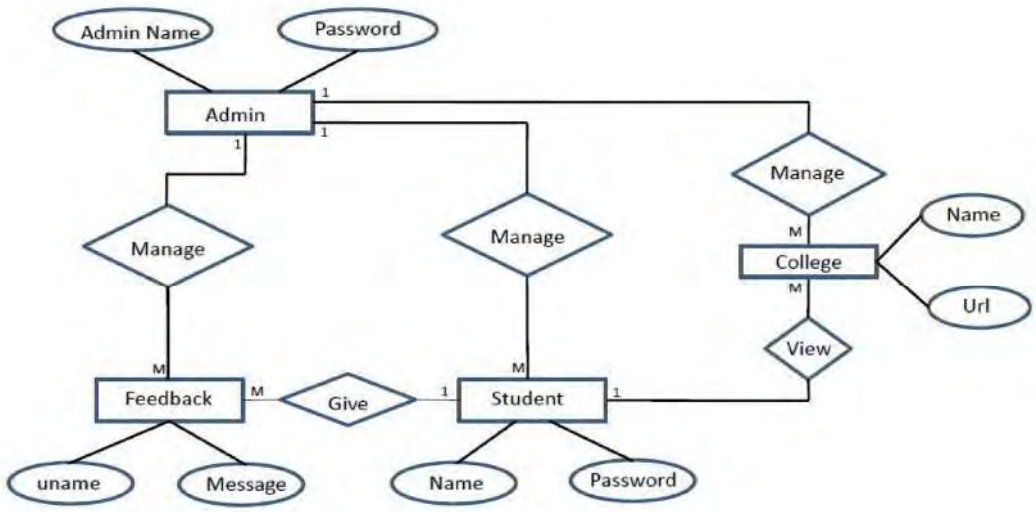
The Department of Graduates Follow-up is responsible for monitoring the affairs of graduates.

It is proposed to be administratively linked to the Principal, where the administration mentors the transfer of students from the academic education environment to the operational environment. The Department also monitors the graduates and learns the news of the university's keenness to maintain a permanent and strong contact with the graduates.

Graduation is the end of the relationship with the university, and try to find jobs for university graduates.

One of the most important objectives of the establishment of the administration is to link the bridges of practical life with the academy and to support continuous communication between the university and its graduate, benefiting the university and its graduates, as well as conducting research studies on academic developments in order to meet the requirements of the labor market. There are now more than one million unemployed, so the technology will be integrated into a database on a website so that information can be accessed in a flexible and easy way





IV. CONCLUSION

This system will be available for general public use through the web interface. A non-registered visitor can look at the list of graduates according to year of graduation or a field of study. He can also look at graduates profiles. The level of profile details shown to the public is limited. By default, a public visitor can only see name and surname of a graduate, year of graduation and a field of study. The faculty endeavors to propagate its graduates.

Therefore graduates can also add some information about themselves into the system during the study such as working experience, knowledge. Graduates can enable to display this information in their profiles for public visitors. Inserted information can be used as an input for generating graduate's curriculum vitae in pdf format, which isprovided automatically. It is in a graduate's competence, which information will be searching pages with their crawlers. A graduate can use it for the building of his virtual web identity on the internet.Our Alumni system solves the problem concerned with graduate's feedback to the faculty with an inquiry module. In this module the faculty can define questions with answers which active graduates can respond. This module should be used for collecting data which are not included in graduate's profiles and have high information value for the faculty

References

http://www.chegg.com/homework-help/questions-and-answers/project-name-alumni-managementsystem-project-project-details-hired-develop-software-track-q20206087

HTTPS://WWW2.DOM.EDU/SITES/DEFAULT/FILES/DOCUMENTS/ABOUT/OIR/ALUMNI_DATA_ENHANCEMENT_PROJECT_FINAL_REPORT_SEPT_12.PDF

Deep Learning Techniques For Detecting Deepfake: A Review

Anushree Deshmukh, Sunil Wankhade
TCET, MUMBAI

Abstract – **Deepfake algorithms can make forged pictures and videos that people can't differentiate them from true ones. The suggestion of technology that locate and prove the truth of virtual visual media is as a result essential. This paper offers a survey of tools and algorithms used to make deep fakes and, additional significantly, methods to locate deep fakes. We present large discussions on challenges, studies developments and guidelines related to deep fake technologies. By reviewing the history of deep fakes and cutting-edge deep fake detection strategies, this gives a comprehensive assessment of deep fake techniques and helps the development of latest and more robust strategies to deal with an increasing number of tough deep fakes.**

Keywords: deep fakes, deep learning, auto encoders, generative adversarial networks (GAN).

I. INTRODUCTION

Deepfake (stemming from "deep learning" and "fake") is a technology which creates fake images or videos of targeted humans by swapping their faces another character saying or doing things that aren't absolutely done by them and humans start believing in such fake, as it isn't always recognizable with the everyday human eye. Deep learning models such as autoencoders and generative adversarial networks have been applied widely in the computer graphics vision to solve various issues [1,2]. These models have been used by deepfake algorithms to examine facial languages and actions of a person and synthesize facial images of another person making equivalent expressions and movements [3]. Deepfake algorithms usually require a large dataset to train models to make realistic images and videos. As public figures such as legislator or celebrities may have a large number of videos and images available online, they are probable targets of deepfake. Deepfakes were used to superimpose faces of legislator or celebrities to bodies in porn images and videos.

The first deepfake video found in 2017 where face of a celebrity was superimposed to that of a porn actor. It is threat to world security when deepfake procedures can be employed to make videos of world leaders with forged speeches for false purposes. Deepfakes therefore can be abused to cause political or religious misunderstanding between countries, to fool public and affect results in election campaigns, or create confusion in financial

markets by creating fake news. It can be even used to

create forged satellite broadcasting images of the Earth to hold items that do not really exist to create chaos in military
There is also advantages of deep fakes such as creating voices of those who have lost theirs voice or updating episodes of web series without reshooting them. However, the number of venomous uses of deep fakes largely dominates that of the positive ones. The growth of advanced deep learning networks and the accessibility of big amount of data have completed the forged images and videos almost unique to people and even to algorithms. The method of creating those forged images and videos is also much easy today as it wants as little as a self-photo or a small video of a targeted individual. Less effort is required to produce an impressively substantial tempered footage. Recent advances can even create a deep fake with just a motionless image. Deep fakes therefore can be a threat affecting not only community figures but also ordinary people.

Finding the fact in digital domain therefore has become gradually critical. It is even more interesting when dealing with deepfakes as they are majorly used to help venomous drives and almost anyone can create deep fakes these days used with present deepfake tools.

Therefore, here have been abundant methods proposed to detect deepfakes. Most of them are also based on deep learning, and thus a battle between malicious and optimistic uses of deep learning methods has been rising.

In Section 2, we present the ideologies of deepfake algorithms and how deep learning has been used to enable such riotous technologies. Section 3 reviews different methods for detecting deepfakes as well as their advantages and disadvantages. We discuss challenges, investigation movements and directions on deepfake recognition as well as interactive program and forensics issues along with research gap in Section 4. Section 5 concludes the paper.

II. DEEPFAKE CREATION

Deepfakes have become popular due to the quality of tampered videos and the easy-to-use skill of their applications to a wide range of users with various computer skills from expert to beginner. These applications are typically developed based on deep learning practices. Deep learning is well known for its ability of instead of complicated and high dimensional data. The first attempt of deepfake creation was Fake App, developed by a Reddit user using autoencoder-decoder blending structure. In that method, the autoencoder extracts hidden features of face images and the decoder is used to reconstruct the face images. To swap faces between source images and target images, there is a need of two encoder decoder sets where each pair is used to

Train on an image set, and the encoder's parameters are joint sandwiched between two network sets.

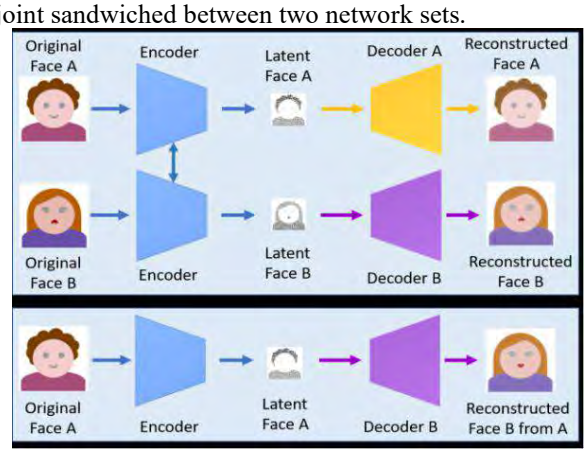


Fig. 1: A deepfake creation model.

This approach enables the common encoder to find and learn the comparison between two sets of face images, which are comparatively unchallenging because faces normally have similar structures such as eyes, nose, mouth places. Fig. 1 demonstrations a deep fake making procedure where the feature set of face A is linked with the decoder B to rebuild face B from the original face A. Table 1 shows some tools for deep fake creation along with its algorithm used and key features

Table 1: Summary of notable deepfake tools

Tools	Algorithm used	Key features,
DFaker	GAN with DSSIM	Reconstruction of face is done by DSSIM loss Function. Keras library-based implementation
Faceswap-GAN	GAN v2.2 with MTCNN	the auto-encoder architecture is provided with adversarial loss and perceptual loss (VGGface) are added to.
Faceswap	GAN with MTCNN and DFL-H128	Parameters of the encoder are shared and pairs of Encoder-decoder.
DeepFaceLab	GAN-er with MTS3FD	Expand from the Faceswap model with new models, Uses multiple face
DeepFake-tf	GAN with DSSIM	Like DFaker but tensor-flow is used for Implementation.

This segment gives a review of deepfake detection methods where we categorize them into two major classes: fake image detection approaches and fake video detection approaches as shown in Fig. 2. The latter is distinguished further: visual artifacts within video frame-based methods and temporal features across frames- based ones. Whereas most of the approaches based on temporal features use deep learning recurrent classification models, the methods use visual artifacts within video frame can be executed by either deep or shallow classifiers.

III. DEEPFAKE DETECTION

Deepfake finding is normally deemed a binary arrangement problem where classifiers are used

between reliable videos and interfered ones. This kind of method requires a large database of real and fake videos to train classification models. The number of fake videos is progressively available, but it is still limited in terms of setting a level for validating many discovery methods.

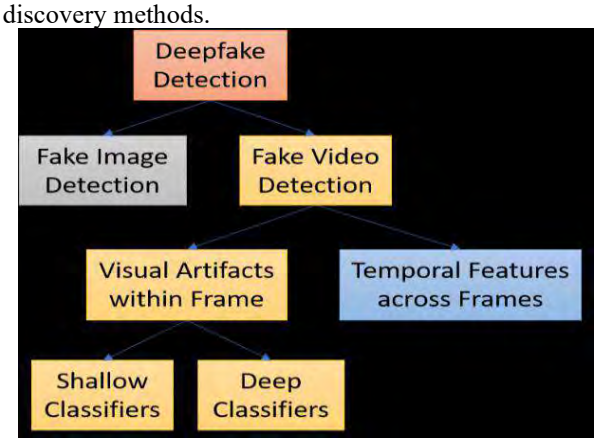


Fig. 2: Classification of deepfake detection.

.

To report this matter, Korshunov and Marcel [6] produced a distinguished deepfake data set containing of 598 videos based on the GAN with the open source code Face swap-GAN. Videos from the openly available VidTIMIT database were used to produce low and high quality deepfake videos, which can efficiently emit the facial features, mouth actions, and eye blinking. These are then used to test various deepfakes. Test outcomes show that the general face recognition systems supported VGG and Facenet [6] are incapable to detect deepfakes successfully. Additional approaches such as lip-syncing and image quality metrics with SVM makes very high error rate when allied to detect deepfake videos from this newly shaped dataset. This increases fears about the serious need of upcoming development of more strong systems that can detect deepfakes from the original.

3.1 Fake Image Detection

Face swap has a few convincing applications in video compositing, transformation in portraits, and especially in individuality safety as it can swap faces in photos by ones from a group of typical images. However, it is one of the approaches that fake attackers enter authentication structures to gain access. The usage of deep learning such as CNN, GAN, SVM, Random forest and multi-layer perceptron has exchanged face images more challenging for forensics as it can reserve pose, facial appearance and light of the photos. Among deep learning produced images, those produced by GAN are possibly most tough to notice as they are genuine and good quality based on GAN's ability to learn supply of the input data and gives novel outcomes with same input distribution.

Maximum works on recognition of GAN produced images do not consider the capability of the recognition models although the development of GAN is continuing, and Several new extensions of GAN are often introduced. Xian et al. [7] used an image preprocessing phase. This rises the pixel level statistical comparison between actual image and false image and needs the forensic classifier to study additional essential and meaningful features, which has better simplification competence than preceding image forensics methods or image stag analysis networks.

Additionally, Agarwal and Varshney [8] company the GAN-based deep fake recognition as a hypothesis challenging issue where a statistical outline was presented by means of the information theoretical study of verification. The least distance among deliveries of genuine images and images produced by a specific GAN is defined, explicitly the vision fault. The logical outcomes demonstrate that this distance rises when the GAN is fewer correct, and in this situation, it is easier to distinguish deep fakes. In case of high-resolution image, a very accurate GAN is required to produce fake images that remain tough to detect.

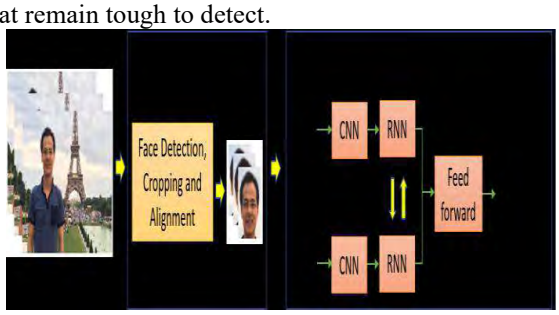


Fig. 3: A two-step process for face manipulation detection.

3.2 Fake Video Detection

Maximum image based deep fake recognition approaches cannot be used for videos because of the robust deprivation of the frame data after audiovisual compression. Also, videos have chronological features that are diverse amongst sets of frames and thus challenging for approaches intended to sense individual fake images. This subcategory emphases on deepfake video recognition approaches and classifies them into two clusters: approaches that uses chronological features and those that explore visual artifacts inside frames.

3.2.1 Temporal Features across Video Frames Based on the opinion that temporal coherence is not imposed efficiently in the procedure of deepfakes, Sabir et al. [10] use of spatio-temporal features of audiovisual streams to sense deepfakes. Video manipulates to carry out on a frame-by-frame basis so that low level objects formed by face manipulations are supposed to further clear themselves as temporal artifacts with irregularities among frames.

A recurrent convolutional model (RCN) was projected based on the combination of the convolutional Network Dense Net and the gated recurrent unit cells to exploit temporal inconsistencies across frames see Fig. 3. The proposed technique is verified on the Face Forensics++ data set, which contains 1,000 videos, and displays promising outcomes.

Guera and Delp [11] detailed that deep fake videos comprise intra-frame conflicts and temporal conflicts among frames. Then they proposed the temporal-aware pipeline technique that uses CNN and long short term memory (LSTM) to detect deep fake videos. CNN is employed to extract frame- level features, which are then fed into the LSTM to create a temporal sequence descriptor.

A Fully-connected network is used afterwards for classifying doctored videos from real ones based on the sequence descriptor (Fig. 4).

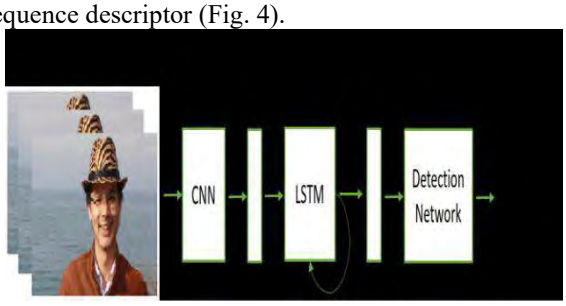


Fig. 4: A deep fake detection using CNN and LSTM.

Similarly, the use of a physical signal, eye blinking, to notice deep fakes was planned in [9] based on the opinion that a individual in deep fakes has a lot fewer frequent blinking than that in original video. A fit adult human would usually blink somewhere between 2 to 10 seconds, and every open and close the eye would take 0.1 to 0.4 seconds. Deepfake procedures, however, frequently use face images existing online for training, which usually demonstrate persons with open eyes. Therefore, without blinking images of persons, deep fake algorithms do not have the ability to produce fake faces that can blink typically. To distinguish actual and false videos, Li et al.

[9] Initially decompose the videos into frames where face areas and then eye region are mined based on six eye signs. After some stages of pre-processing such as aligning faces, extraction and scaling the bounding boxes of eye sign points to generate new arrangements of frames, these cropped eye area arrangements are distributed into long-term recurrent convolutional networks (LRCN) for dynamic state prediction. The LRCN contains of a feature extractor built on CNN, a sequence learning built on long short-term memory (LSTM), and a state prediction built on a fully connected layer to forecast probability of eye open and close state. The eye blinking illustrates robust temporal dependencies and therefore the application of LSTM helps to capture these time-based patterns efficiently. The blinking rate is calculated built on the predicted outcomes where a blink is defined as a highest above the threshold of 0.5 with interval less than 7 frames. This technique is assessed on a dataset collected from the

internet containing of 49 videos and their matching fake videos produced by the deep fake algorithms.

3.2.2 Visual Artifacts within Video Frame

As observed in the preceding subsection, the approaches using temporal patterns across video frames are typically based on deep recurrent network models to sense deep fake videos. This subsection inspects the additional method that usually decomposes videos into frames and explores visual artifacts within single frames to get distinguished features. These features are then dispersed into either a deep or shallow classifier to discriminate between false and true videos. We thus assemble approaches in this subsection based on the categories of classifiers, i.e. either deep shallow.

3.2.2.1 Deep classifiers

Deepfake videos are usually formed with few resolutions, which need a face warping method (i.e., rotation and shear) to matches the original ones. Because of the resolution contradiction between the warped face part and the nearby context, this procedure leaves objects that can be noticed by CNN methods such as VGG16, ResNet50, ResNet101 and ResNet152. A deep learning technique to spot deepfakes based on the objects observed throughout the face warping step of the deepfake generation algorithms was proposed in [12]. The proposed structure is assessed on two deepfake datasets, i.e. the UADFV and Deepfake TIMIT. The UADFV dataset [15] contains 49 actual videos and 49 forged videos with 32752 frames in total. The Deepfake TIMIT dataset contains bad quality videos of size 64*64 and another good quality videos of 128 x 128 with 10537 unique images and 34,023 forged images mined from 320 videos for each set. Performance of this technique is associated with other predominant approaches such as the face tampering recognition technique two-stream NN, Head Pose [15], and two deepfake recognition MesoNet methods, i.e. Meso-4 and MesoInception-4 [13]. Advantage of the proposed technique is that it need not to produce deepfake videos as forged examples before training the detection methods. As an alternative, the false examples are produced dynamically by mining the face area of the unique image and aligning it into several scales before applying Gaussian blur to a scaled image of chance choice and warping back to the unique image. This decreases a huge time and computational resources associated to other approaches, which need deepfakes are produced in advance. Recently, Nguyen et al. [14] proposed the usage of capsule networks for distinguishing doctored images and videos. The deepfake face swap dataset created by Afchar et al. [13].

IV. RESEARCH ISSUES

Deepfakes are gradually damaging to confidentiality, society safety and democracy. Diverse approaches for detecting deepfakes are proposed by many researchers as soon as this menace was familiarized. Initial efforts were based on handcrafted features gained from objects and on social context in which deepfake is revealed. This is serious as deepfakes are getting more photorealistic and it is extremely projected that recognition software will be lagging behind deepfake creation techniques.

4.1 Limitations of Existing system

In the technique where the recognition is done by eye blinking technique [9] the dataset used was very small thus system accuracy will be reduced. The experimental outcomes show great performance of the planned technique in detecting false videos, which can be further enhanced by seeing dynamic pattern of blinking. Detection is done by frame by frame basis so there might be chances of inconsistency [10]. RNN gives better result than CNN [11][12][13].

4.2 Effects of Deepfake

Misinformation: Folks are more probable to have a response to misinformation in the form of forged image, audio, and video content, which allows the doctored media to spread more rapidly than purely written fake data. Additional, images and video have been recommended to generate a Mandela effect, the formation of memories that never happened.

Exhaustion of serious thinking: It will take more effort for people to determine whether data is true, especially when it does not originate from reliable actors. Uncertainty around content reliability might also put off an distinct from sharing correct content, dropping the distribution of true information.

The deceiver's dividend: The presence of entirely artificial content offers an avenue for actors to deflect charges of impropriety based on footages and video, by appealing the source material has been faked.

These results are troubling and will be most persistent in the future, as deep fake excellence increases and social alertness lags

V. CONCLUSION

Deep fakes have started to wear down trust of people in media contents as seeing them is no longer corresponding with believing in them. This is critical nowadays as the technologies for creating deep fakes are progressively approachable and social media platforms can spread those fake contents rapidly. This paper has revised the state-of-the-art methods and a summary of typical approaches are provided. It is noticeable that a fight between those who use progressive machine learning to create deep fakes with those who make effort to detect deep fakes is rising. Deep fakes' good quality has been increasing and the Performance of discovery methods needs to be enhanced consequently. Detection methods are still in their early stage and various methods have been suggested and evaluated but using fragmented data sets. An approach to progress performance of detection methods is to create a rising updated standard data set of deepfakes to authorize the

ongoing development of discovery methods. This will simplify the training process of discovery models, particularly those based on deep learning, which involves a large training set. The current detection methods mostly focus on disadvantages of the deepfake generation pipelines, i.e. finding weakness of the participants to attack them. This kind of information and knowledge is not always available in adversarial surroundings where attackers usually attempt not to reveal such deepfake creation technologies. This is a real challenge for detection method development and a future research needs to focus on introducing more robust, scalable and generalizable methods.

References

[1]. Tewari, A., Zollhoefer, M., Bernard, F., Garrido, P., Kim, H., Perez, P., and Theobalt, C. (2018). High- fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. DOI: 10.1109/TPAMI.2018.2876842.

[2]. Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X.,Huang, X., and Metaxas, D. N. (2019). StackGAN++: Realistic image synthesis with stacked generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8), 1947-1962.

[3]. Lyu, S. (2018, August 29). Detecting deepfake videos in the blink of an eye. Retrieved from http://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072

[4]. Damiani, J. (2019, September 3). A voice deepfake was used to scam a CEO out of $243,000. Retrieved fromhttps://www.forbes.com/sites/jessedamiani/2019/03 avoice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

[5]. Chorowski, J., Weiss, R. J., Bengio, S., and Oord, A. V. (2019). Unsupervised speech representa- tion learning using wavenet autoencoders. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 27(12), pp. 2041-2053

[6]. Korshunov, P., and Marcel, S. (2019). Vulnerability assessment and detection of deepfake videos. In The 12th IAPR International Conference on Biometrics (ICB), pp.1-6.

[7]. Xuan, X., Peng, B., Dong, J., and Wang, W. (2019). On the generalization of GAN image forensics. arXiv preprint arXiv:1902.11153.

[8]. Agarwal, S., and Varshney, L. R. (2019). Limits of deepfake detection: A robust estimation viewpoint. arXiv preprint arXiv:1905.03493.

[9]. Information Forensics and Security (WIFS) (pp. 1-7). IEEE.

[10]. Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., and Natarajan, P. (2019). Recurrent convolutional strategies for face manipulation detection in videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 80-87).

[11]. Guera, D., and Delp, E. J. (2018, November). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE

[12]. Li, Y., and Lyu, S. (2019). Exposing deepfake videos by detecting face warping artifacts. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 46-52).

[13]. **13)**Afchar, D., Nozick, V., Yamagishi, J., and Echizen, I. (2018, December). MesoNet: a compact facial video forgery detection network. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-7). IEEE

[14]. **14)**Nguyen, H. H., Yamagishi, J., and Echizen, I. (2019, May). Capsule-forensics: Using capsule networks to detect forged images and videos. In 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2307-2311). IEEE.

[15]. Yang, X., Li, Y., and Lyu, S. (2019, May). Exposing deep fakes using inconsistent head poses. In 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8261-8265). IEEE.

[16]. Matern, F., Riess, C., and Stamminger, MAIN. (2019, January). Exploiting visual artifacts to expose deepfakes and face manipulations. In 2019 IEEE Winter Applications of Computer Vision Workshops (ICCVW) (pp. 83-92). IEEE

[17]. Koopman, M., Rodriguez, A. M., and Geradts, Z.

[18]. (2018). Detection of deepfake video manipulation. In The 20th Irish Machine Vision and Image Processing Conference (IMVIP) (pp. 133-136).

[19]. Hsu, C. C., Zhuang, Y. X., and Lee, C. Y. (2019). Deep fake image detection based on pairwise learning. Preprints, 2019050013.

[20]. Zhang, Y., Zheng, L., & Thing, V. L. L. (2017). Automated face swapping and its detection. 2017 IEEE 2nd International Conference on Signal and Image Processing(ICSIP).doi:10.1109/siprocess.2017.8124497

Review of an Indoor Navigation System Using NFC Technology

Heeth jain ,pratik gupta,pritesh jain,ravindra chabhadia ,shubham gadia
TCET, Mumbai

Abstract— **There are many problems faced in the existing navigation system. Here we present a NFC (Near Field Communication) based indoor navigation system called NFC Internal in order to improve and enhance the current Indoor Navigation System. NFC Internal enables an easy data transfer for indoor navigation systems just by touching tags spread over a building or a complex. This review has some pros and cons but is definitely improved then previous systems. The system requirements were discussed and necessary upgradation were done to the previous technology.**

Keywords- NFC, NFC Internal, Indoor Navigation

I. INTRODUCTION

Navigation is the thing that helps one to reach from one point to another. While one is traveling on an intended route, navigation systems provide reading, controlling and updating the movements of one's position and orient. The demands and usage of outdoor navigation systems are increasing incredibly today as all people use PDAs, mobile devices or personal navigation assistant and navigation systems can run on those devices.

Global Positioning System (GPS) is generally the base of outdoor navigation systems, it provides reliable location information in almost all weather conditions and at all times on or near Earth as it is a space based global navigation satellite system. It is a well explored and standardized area of research. But due to the absence of line of sight to the satellites the GPS receivers cannot perform well indoor environments. So because of the unavailability of GPS indoor environment, the indoor navigation system has become a recent research area. In order to circumvent the lack of excellence, variety of technologies are tested and new designs are generated for indoor navigation.

For indoor navigation systems the existing solutions are grouped as network based navigation system that are based on networking technologies such as sensor networks[4], and an Independent navigation system providing autonomous user position. Technologies such as Bluetooth, Ultra Wide Band (UWB), WiFi or Radio Frequency Identification (RFID) are used by network based navigation systems.Wifi and UWB technologies provide higher position accuracy than Bluetooth and RFID technologies.

The position accuracy of a Bluetooth depends on the amount of cells used, it is a simple, compatible short range Communication technology. The same for RFID technology depends on the type of tags, which can be either active / passive, and also the amount of these tags. The RFID based indoor navigation solutions up till now are generally based on usage active RFID tags and require extensive usage of active RFID tags to get good position accuracy. Opposite to the passive tags, to increase the transmitting distance active tags contain embedded batteries. The high cost of the active tags is the major drawback of using active RFID tag based solutions. Studies in this area also indicate that it does not provide an efficient tracking system. WiFi and UWB technologies have their own limitations as well. WiFi requires expensive access point in any area where the person needs to be tracked. Due to some technical problems of technology (e.g. antenna mismatch, low power emission, and possible external interference from other systems) an efficient indoor navigation system cannot be ensured in case of UWB. Independent navigation systems based on dead reckoning (DR) methods are the other popular systems. There are 2 types of positioning defined when marking position of a person on a map. I.) By making use of help from enough number of assisting devices such as satellites, the FIX position determines the location. I.) Whereas the current position is determined using the last fix position, the speed of the item, the route and the time required to calculate the last fix position.
This methodology is used by navigation systems based on DR methods. They use Micro Electro-Mechanical Sensors (MEMS) that include electronic barometers, magnetometer and accelerometers. When the calculation of new position is based on previous DR position, the errors of the estimation process are cumulative. The performance of the system is affected by the large errors, so this acts as a major drawback of DR method based navigation system.

Assisted GPS (A-GPS) systems which enlarge the working area of GPS technology is another independent, existing technology. With A-GPS systems, through an assistance data server the indoor GPS. signals are processed efficiently. The assistance data server is connected to reference receiver. However, the strength of the signal indoors is sometimes too

low. The high installation cost and the complexity of the system design are the major drawbacks of the indoor navigation system. Also, the other existing positioning systems are far from providing an accurate position in large buildings [4],[5]. Therefore inside large buildings consisting of many rooms, floors and large halls there is a strong requirement of an efficient and low cost indoor navigation system. At any time the navigation systems for indoor environments are desirable.

For some instance, a person is entering a building for the first time and he may want to go to an office inside the building. Such cases are very common at University, airports, complexes, hospitals, and shopping centers etc. In such cases, a user friendly navigation system guiding people through huge buildings consisting of thousands of rooms can be helpful. This idea presents an innovation of Low cost indoor navigation system called NFC communication technology.

The main idea is to orient users by NFC enabled mobile phones, which also have an embedded indoor navigation application. While this application is oriented by the user by gathering destination point. The mobile devices gathers the current position from NFC tags and shares the coordinate data with the application. The user can now determine the current position inside the building by a simple touch on the device to tags which are spread inside a building.

The remainder of this paper is organized as follows:

In section 2, solutions to indoor navigation based on NFC are presented together with the system design and it's implementation. Section 3 consists of the conclusion of paper as well as highlights the future work.

II. NFC INTERNAL

NFC is bidirectional short range, wireless communication technology. This communication occurs between 2 devices within the range of few centimeters. 13.56MHz signal with a bandwidth not more than 424kbit/s is used. This technology is based on Radio Frequency Identification (RFID) technology. It can be operated on acrd emulation, reader/writer. It is also used in peer to peer operation modes where communication occurs between a mobile phone on one side and an NFC reader, a passive RFID tag (NFC tag), or a mobile phone on the other side. It is very simple and easy to use technology. The user only needs to carry and use an NFC enabled mobile device.

Indoor Navigation application must be OTA(over the air) preloaded to the smart card. The user just needs to simply touch the URL tag that contains the URL of the indoor map information just before entering the area or building. The map present on the site is OTA downloaded to users mobile from Mapserver . This map information is afterwards used by the device as the Indoor Navigation application starts automatically. So as the person selects destination of his voyage inside the building, the indoor navigation application provided the most optimized route to the destination. As the user navigates through the halls, they have the option to touch to the Reference Tags (NFC tags) spread over the building to fix their current positionon the map, & then get instructions to reorient their position to destination or to create a new optimal route to the destination is necessary. For indoor navigation systems, NFC technology is a coherent solution then all the current ones. The use of NFC can be extended to Internal and more other cases. The existing outdoor navigation systems and the NFC Internal can be combined and such kind solution provides higher functionality.
For example, if a new student is currently staying in their home and has to meet his advisor for some work, but they don't not know where the office of their advisor is in the campus. In this case, user will need an mobile device with NFC enabled which has a GPS-based mobile navigation system for outdoors as well as indoor navigation application. The student starts the process by entering the name of the university campus and name of the advisor. GPSbased mobile navigation application will help the student until they reache the respective university campus.

Map of the surrounding area is shown on the mobile device screen, which the help of student’s current position and a route to the campus are also indicated as well [1]. As student moves to the destination, the mapped area and the student’s position are dynamically updated on the map. GPS based mobile navigation systems can also provide audio guidance to the student . When student reaches to the university campus, they need to simply touch the URL Tag on the entrance in order to get indoor map information and navigate inside.

After the map is downloaded from MapServer, the outdoor application quickly shares the destination information with the indoor application on the mobile device which is a seamless solution. So, the student does not need to enter the destination address again. Now, user can start to navigate inside the campus by touching to the Reference Tags as described above.

II System Design

Spatial Information Maps:

Current Indoor map database is generally based on 2D graphical representations which are developed by CAD (Computer-Aided Design) systems [3].

In order to [7], the CAD drawings of an indoor environment can be separated into floor maps of each floor in the building, and each CAD drawing of floor

plan can further be converted into separate spatial information maps that annotate structural features such as walls, doorways, elevators, and staircases. In each separate spatial information map, the reference points are also expressed in form of vector spatial data structure which uses the 2D Cartesian coordinate system and each map is stored in the Map Server for the Indoor Environment with a unique floor identifier, and also a unique building identifier if the indoor is a complex or campus.

Reference Tags for NFC Internal:

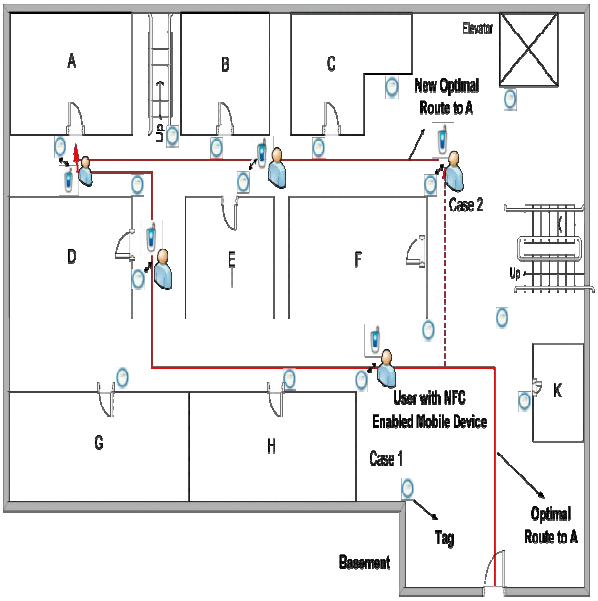


Fig. 2. Navigating to Destination

There is a possibility that the user might get out of the NFC technology in indoor navigation systems has a potential

Vector spatial data structure is used instead of raster data structure. Use of vector spatial data allows efficient encoding topology and also network linkages can be efficiently employed [9]. Thus it is more useful for accurate positioning.

B. How NFC Internal Works

This section shows how NFC Internal works in detail. The NFC Internal is comprised of two phases: initiating the indoor navigation system and navigating to destination. As described before, the main problem starts with the user who needs to reach to an intended point, but does not know the exact location of the place. In such situation, the user who has an NFC Internal system on her an NFC enabled mobile device is only required to follow the phases hereunder.

1) Initiating NFC Internal:
1. As seen in Figure 1, the user touches to the URL Tag which contains the URL of the indoor's map information on the MapServer. This tag is placed on the entrance of the building. The NFC enabled mobile device gets the address as it touches to the URL Tag.

2. Mobile device connects to the URL via OTA and requests the map information from the map Server, and the map information is loaded to the mobile device afterwards.
3. After loading the map, indoor navigation application on the smart card automatically starts and converts the map data into link-node model as 2D network with topological relationships.

4. The application asks the user to enter the destination point.
User specifies destination point just by choosing the person's name.
5. The application quickly computes the best route by using
Dijkstra's shortest path algorithm.

2) Navigating to Destination:
After the route is computed, the application starts to orient the user towards the destination. As the user navigates along the path, she can touch to any Reference Tag to validate her navigation. User is only required to touch to a tag on her way to get this valuable help. The location data on the Reference Tag is transferred to indoor navigation application at this moment. It is obvious that information on the tag also shows the user's current location. The application uses this information to check whether the user is on the intended route or not. The application forms simple plain instructions as forward, left, right, backward etc. to orient the user which can be easily followed by the users. Also there are instructions to use stairs upward / downward or to use elevator up / down to a specific floor. These instructions are displayed on the screen of the mobile device.
Let's concentrate on the simple case shown in Figure 2. According to the scenario, user wants to be directed to Office A, and starts navigating inside the building for this purpose. She touches to the first Reference Tag that she sees on her way, after which the location data on tag is transferred to the application on mobile device (Step 1'). So, the application figures out that she is on the intended route and gives user the first instruction as: "go straight ahead for 25 meters, turn right". Similarly user touches to another Reference Tag on her route

and application finds out that user is still on the correct
route, so gives instructions to the user (Step 2').
route, towards Office A for example, and touches to a Reference Tag which is not on her route (Step 1"). With the transfer of location data, the application figures out that the user is out of route and calculates a new shortest path to Office A and gives user a new instruction as: “go straight ahead for 50 meters”. Similarly user touches to another Reference Tag on her new shortest route to Office A and application figures out that user is on the intended route indeed, and gives instructions to the us (Step 2"). As the user arrives the destination point, Office a (Step 3), the application tells user that she reached the destination, and ends the ongoing process.

III. CONCLUSION

This paper presents a new, reliable, and seamless indoor navigation solution that helps to create smart and contextaware environments. The use of NFC technology in indoor navigation systems has a potential to increase the usability of these systems. Comparing to other existing indoor navigation solutions, NFC Internal has many advantages hereunder:

• Reduces the cost of indoor navigation systems by using cheap passive tags,
• Minimizes response time, because the time required to transfer data from NFC tag to mobile device and the time required to generate application’s new path is little,
• Provides accurate position and orientation information, so the orientation of the user to the destination is facilitated,
• Eliminates the need for a server or a terminal to orient position, so location privacy of the user guaranteed,
• Provides exclusive control over her location data for the user. On the other hand there is a limitation of the proposed system.

In the system, on the go position information cannot be provided,
since user can learn her position information as she touches an
NFC tag.

Overall, we think that the proposed system is very simple to use and has several benefits to users.

IV.References:

T. Ishikawa, H. Fujiwara, O. Imai, and A. Okabe, “Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience,” Journal of Environmental Psychology, vol. 28, pp.
74-82, March 2008.

[1] Wikipedia, Global Positioning System, Available: http://en.wikipedia.org/wiki/Global_Positioning_System. Access date:
10 October 2010.

[2] P.-Y. Gilliéron, and B. Merminod, “Personal navigation system for indoor applications,” in Proc. 11th IAIN World Congress, Berlin, 2003.

[3] V. Renaudin, O. Yalak, P. Tomé, and B. Merminod, “Indoor navigation of emergency agents,” European Journal of Navigation, vol.
5, pp. 3645, July 2007.

[4] R. Ivanov, “Indoor navigation system for visually impaired,” in Proc.
Int. Conf. on Computer Systems and Technologies, Sofia, 2010, pp.
143149.

[5] P. Ruppel, and F. Gschwandtner, “Spontaneous and privacy-friendly mobile indoor routing and navigation,” in Proc. 2nd Workshop on Services, Platforms, Innovations and Research for New Infrastructures in Telecommunications, Lübeck, 2009, pp. 2574-2583.

[6] A. K. L. Miu, “Design and Implementation of an Indoor Mobile Navigation System,” M.S. thesis, Dept. Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA, 2002.

[7] P.-Y. Gilliéron, D. Büchel, I. Spassov, and B. Merminod, “Indoor Navigation Performance Analysis,” in Proc. 8th European Navigation Conference GNSS, Rotterdam, 2004.

[8] I. Heywood, S. Cornelius, and S. Carver, An Introduction to Geographical Information Systems. UK: Pearson Prentice Hall, 2006, ch. 3.

[9] NFC Forum, Available: http://www.nfc-forum.org. Access date: 10
October 2010.

Secure Data Aggregation and data Transmission using HMAC Protocol in Cluster base Wireless Sensor Network

Sunny Sall[1], Dr. Rajesh Bansode[2]
Department of Computer Engineering St. John College of Engineering and Management[1], Department of Information Technology, Thakur College of Engineering and Technology, Maharashtra, India[2]
sunny_sall@yahoo.co.in ,rajesh.bansode@thakureducation.org

Abstract— **Wireless Sensor network (WSN) facing many issues in due implementation in vulnerable environments. Various researchers has already defined a systems for data transmission with WSN, still such system having high data loss issues, maximum packet delay as well as packet overhead during the data transmission. Different parameters has to consider to improve such losses like cluster network generation, data aggregation, secure data encryption and data transmission using distribution approach produces effective outcome and eliminate such issues. Moreover implementation with cluster network and selection a cluster head (CH) based on trust which produces much effective results and provides flexibility to system. This is research we proposed secure data transmission in cluster network and investigate the Quality of Service (QoS) parameters of entire execution. Initially we create different clusters with collaboration of multi sensor nodes while each node consist individual battery power as energy. According to highest energy we calculate the trust and such node defined as CH, Data Aggregation (DA) and Broadcast Tree Construction (BTC) two different techniques have been used for eliminate network lifetime or cut generation in system. In partial experiment analysis system shows improve QoS parameters like Throughput, delay, packet overhead etc respectively. It also enhance the network life because of proposed energy conservation approach.**

Keywords - Cluster Network, WSN, Data aggregation, Broadcast Tree Construction, QoS.

I. INTRODUCTION

In this era, the dagger development of mobile computing expedients that primarily embody laptops, personal digital assistants (PDAs), as well as hand-held digital devices, is driven by a revolutionary change in the computing world. Security approaches to eliminate different network attacks like wormhole attack [1] in untrusted environment. To identify intruder or threats from large traffic defined in [2]. Computing will not simply place trust within the power provided by private computers, and the idea of gift computing often arises and becomes one within the applied science society at every research hotspots [3]. Throughout this environments a path behalf of the two hosts might contains steps through one or extra nodes inside the painter. an important draw back in associate degree passing mobile ad-hoc network is finding and maintaining routes since host quality can

cause topology changes [4]. Several routing algorithms for MANETs are proposed inside the writings which they differ inside the painter. New routes square measure set up and existing ones square measure modified. Basically MANET networks square portion extra susceptible to suffer from the hateful performances. This paper we proposed an wormhole attack detection and prevention approach using secure mechanism of detection such malicious behaviour. This paper we proposed an various attack detection and prevention approach using secure mechanism of detection in malicious behaviour [5] and investigate the proposed experimental analysis.

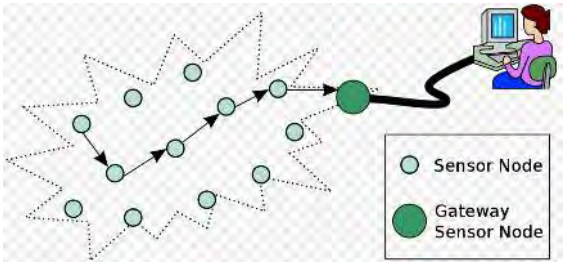


Figure 1 : Representation of a wireless sensor network [6]

The above figure 1 shows basic WSN view in sensor network, which is basically similar to wireless ad hoc networks in the intelligence that they rely on wireless connectivity and unstructured establishment of networks so that sensor data can be ecstatic wirelessly.

RELATED WORK

Yashpal sinh Gohil et. al. [1] proposed a true Link verifies whether or not there's an immediate Use truelink to link a node to its adjacent neighbor. The main disadvantage is that TrueLink only works on IEEE 802.11 devices that are backward compatible with an update to the computer code. An approach to trip time (RTT) is emerging to beat the problems of further hardware exploitation. The RTT is that the time taken to deliver RREQ to a supply node and to collect RREP from endpoint. A node should calculate the RTT between its neighboring nodes and itself. The malicious nodes are priced higher for RTT than different nodes. This method will determine the supply of its real and misbehavior neighbors. This detection technique is efficient only in the case of hidden attacks.

Vikaskumar Upadhyay and Rajesh Shukla [2] describes Numerous methods are planned to be used to identify wormhole attacks using a leash method for the packets. Packet leash is the technique that guards against wormhole attack. The leashes are either spatially or temporally categorized. All nodes inside the network should have data from their own position and stable synchronized clock in spatial leashes. This

requires its own current location when a sender sends the data packet and UTC accepts wormhole nodes. Directional data are exchanged between source and destination during this technique. Destination by scrutiny will find the wormhole the Received signal from malicious nodes and gave directional info. The relation to the wormhole is observed if the supply signals and intermediate nodes are distinct.

Amit Kumar and Sayar Singh Shekhawat [3] describes, in Authors have criteria of contact that are primarily focused on a paradigm of research to ensure safe communication. Another such assault is a wormhole attack, in which two or many nodes are often flustered together to gain knowledge calculation and communication. A wormhole compromised network is illustrated during this post, and the research model can be used for secure communication within the network being targeted. The network model is designed to identify the secure communication node, so that connectivity is maximized. Additionally, in mostly mobile network based attack, the writers address a way of generating the safe path. The model eventually provided the optimized parameter for reconciliation communication. Results show the improved performance of model in terms of the communication throughput and reduced the loss.

Ashish Kumar Jain and Ravindra Verma [4] authors To search for aggressive or legitimate nodes inside the network, build calculative confidence values for the nodes. Authors propose a way to defend against victimizing wormhole attack combination of parameters such as energy, connection range and node buffer length. A node's confidence value is calculated backed by these parameters. This node confidence value is then compared with the threshold value of network faith. This comparison was supported and one can find out if The selected node is either malicious or valid, or is not. The planned methodology consists of two phases: first, the analysis of the network parameter and the calculation of the threshold, and second, the implementation of the protection on the routing protocol that prevails. Results area unit assessed by expanding the AODV protocol as well as evaluating the efficiency of the proposed routing protocol and matching it to the insecure AODV. The performance of the proposed confidence-based approach, based primarily on defense, is cost-effective and reasonable against comparative findings wormhole attacks in MANET.

Rajan Patel Anal Patel, Nimisha Patel [5] this Survey is The different techniques used to identify authors of wormhole attacks suggest an approach for the identification and prevention of wormhole attacks. A projected defensive approach against wormhole attack based primarily on the Hash-based Compression Function (HCF) that is really mistreatment of any secure hash operates to chip a hash field for the RREQ packet and the projected approach looks terribly promising compared to other proposed solutions in literature.

Dhruva Patel et. al. [7] system describes various safety hazards which affect different levels of safety. Because of its simplicity and self-routing nature, MANET is responsive to various threats, and various attacks can breach separate layers within the network. There are numerous attacks and each attack has its own impact on different layers since some can only affect a single layer of the network while others can hit other layers, i.e. depending on the nature of the intrusion on how it reacts. Wormhole assault is now a network layer intrusion capable of crippling the communication network entirely. all the attacks.

Juhi Biswas et. al. [8] system present AODV With the implementation of the MANET and Wormhole Attack Detection and Prevention (WADP) technique on this revamped AODV, the routing protocol is modified to detect and prevent wormhole attacks in real world. So that fraudulent nodes are found within the network and the false positive limitation node authentication function is eliminated. Simulation tests also indicate that node authentication does not eradicate false positives although it helps to chart the actual location of the hole jointly and could be a relatively double test for hole attack detection. This algorithmic rule doesn't use any special hardware for police work wormhole attack.

Kapil Raghuwanshi et. al. [9] take Initiative to eliminate or minimize the impact of a hole assault by offering a solution within the initial route setup point that could sense the presence of a void. Such resolution is based on the hop count analysis approach, i.e. hop count is used as a parameter for characteristic shapes involving tunnel space. Hop counts are evaluated to detect hostile nodes. Simulation of the planned work is completed in many node and traffic situations in the presence of a wormhole attack. The simulation results The predicted strategy demonstrates superior performance as PDR and attendance decreases however, in fact, "normal end-to-end gap" rises. It has been noticed within the analyzed state of affairs that the MAODV has a superior performance than AODV. Modified AODV is ideal for wormhole attack identification and bar detection. This strengthens the insecure circumstances of the packet distribution level, with a borderline reduction in turnout and a reasonable raise in end- to-end delay.

System [10] Throw light on a weak wormhole attack, during which two or more malicious nodes type a tunnel as a relay mechanism into the packets themselves. Such an attack may result in the packets sent being selectively diverted, fabricated and changed. During this paper AN identity is planned on the side of the clusters to shield the network from hole attacks based primarily on the signature theme. The designed theme need not transmit any credentials between nodes, thereby overhead calculations. Cluster During which cluster heads area unit selected in such a way that they cannot be harmful, specifically focused architecture is employed. This style works in 3 steps. The effects of the simulation indicate the improved performance of the theme in

terms of throughput, packet delivery ratio and end-to-end delay. According to [11] proposed protocol is based on Neighbourhood overhearing and frequent analysis of key and different tables and expected procedure information was found to be safe and a few region assault units are checked on it. Wormhole assault is observed by nodes being overheard. The results show that M-AODV has been strengthened in terms of the packet transmission magnitude relationship and the latency has therefore been lowered further, however the overhead amount has been raised. M-AODV jointly improves network reliability and stability. Thus, the planned protocol is proven to be protected in simulations alongside wormhole and blachole attacks.

Luo It. Al. [15] Fixes problems with WSN because every node is created to handle all the information from a forward-looking and network management-friendly, autonomous system to all levels, from the appliance layer to the appliance layer. This works very well, especially thanks to the algorithms developed with the short-shortcut WSN, they do not create simplicity and completeness when trying to implement widely and widely distributed different and low-energy WSNs at the same time.

Gante et al. [16] for example announce good organization of SDWSNs to boost potency and Overcome the inherent difficulties of common WSNs. The management theme is predicated on a Base Station Design for WSNs with Associated Integrated Controllers. The controller creates forwarding rules that area units stored in flow tables from location knowledge acquired through the localization technique processed in the layer layer application layer of the design.

Olivier et. al. [17] projected gradable design referred to as software package outlined Clustered detector It is believed that multiple base stations network (SDCSN) are used as controllers which additionally play the role of cluster head. Clusters of large nodes are divided into clusters, and each has a cluster head. The cluster head controls and coordinates the detector nodes in each cluster, and the knowledge processed in each cluster is diverted to the cluster head.

Oliveira et al. [18] style associated implement associate design supported Even within the framework many controllers in WSN referred to as Tiny-SDN supports Tiny-OS with a style structure consisting of SDN-enabled detector nodes and SDN controller node. This fixes issues such as in-band management, high communication delays and restricted power offers.

Tootoonchian et. a. OpenTM [19] projected associate very Put a little effort into this technique or work for new techniques observation SDN-based WSNs associated remains an exposed analysis space. Freshly but, work has been conferred on a network mensuration design supported SDN for observation of WSN data like routing path per- packet, the ratio each Link and hence the delay in each ointment-hop configuration.

East-West Apis modify supervisors that area unit within the similar dominion or neighbouring dominions to express with one another [20]. it's vital to say here that SDN isn't concerning SDNs have been shown to have the ability to change network management and allow innovation through network programmability, rather than enhancing network performance.

The challenges to think about once integration a quality management theme contains management the impact of nodes coming In the network and therefore the nodes perform network functions, implement functions, and different network attributes on QoS. A number of solutions and process steps have been provided by Zhou et al to stop the problems associated with the SDN controller nodes and offer WSN efforts. [21].

II. PROPOSED SYSTEM DEISGN

Figure 1 describes the cluster creation process and transmission between source and destination node. At the end of each TS network nodes verifies sensed data and broadcast messages to nodes within given Cluster Distance (CD) for cluster creation. Cluster creation uses the Relay Node (RN) and CD to group the sensor in same cluster. Upon accepting the broadcasted message each node verifies the value of RN. If its value is within RN it stores in its memory and compares CD with each node's distance. If the distance between nodes is same or less to CD and sensed value is within given RN then those group of nodes forms a cluster. The nodes NID which are related they will not broadcast message for cluster creation. Nodes which are not participate in cluster creation process based on *RN* and *CD*.

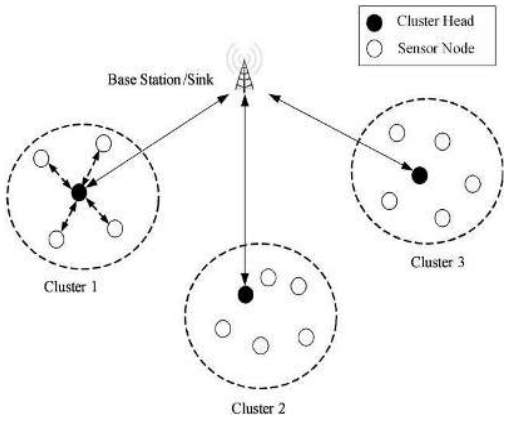


Figure 2: Proposed system cluster scenario

III. CLUSTER HEAD SELECTION :

Input: Cluster set with nodes.
Output: Ch selection with remaining sensor node.
Step 1: select all nodes as initial population.
Step 2: Select evaluation set
Step 3: Apply crossover on similar power nodes. **Step 4:** Apply mutation on each sensor node. **Step 5:** Apply fitness on all nodes power
Step 6: select best node using rout let wheel selection.
Step 7: Check GA evaluations

Step 8: Select final max energy node as CH node.

IV. DATA TRANSMISSION USING BTC

A dynamic cluster it will be designed when the goal reaches several cluster boundaries. A difficult role problem is how the device identifies the situation, particularly in a fully distributed environment, when the goal reaches the boundaries. We use boundary nodes to solve this problem in a fully distributed way.

The following assumptions are taken in order to design the proposed algorithm.

In this work we have carried out an algorithm which is wormhole attack identification has been through in a cluster based network creation manner to eliminate the wormhole attacks. Basically AODV routing protocol is utilized as the fundamental network topology. A multi-layer method is used to determine if a node is participating in a wormhole attack. Introduces the layered approach to decrease the capacity of dispensation on respective cluster heads. Due to safety point of view, this also decreases the risk of a cluster head should be hacked by attacker.
The complete network is separated in clusters sets illustrated in figure 1. Sometimes clusters might be corresponded or separate. Every cluster contains a single cluster head as well as number of cluster member nodes. Member nodes forward on the data only to the cluster head when any nodes want to send data to Cluster Head (CH). The CH is responsible for forward on the collective data to all its other cluster members. The CH is selected enthusiastically and preserves the routing information.

Input: Primary source node *Sender_node*, Destination node *Dest_node*, Group of nearest nodes Neigh_node *[],* node id as *N_id,* node energy *N_eng;*
Output: From source to destination way based on the given algorithm.

Step 1: initially system select the *Sender_node* and *Dest_node* on dynamically

Step 2: select the packet or file f for info broadcast.

Step 3: if (*file or data* ! =null)

Step 4: read each byte *bytes* form *file or data* when reach null

Step 5: send data, initialize *cost_filed_1, cost_filed_2, parent_filed_1, parent_fileld_2*

Step 6: while (nd[i] when reach NULL)

cost_filed_1=node[i]_eng parent_filed_1= node[i]_id

cost_filed_2 =node[i+1].eng parent_filed_2= node[i+1] _id

Step 7: if (*cost_filed_1*> *cost_filed_2*) *cost_filed_2*=null *parent_filed_2*=null

Else

parent_filed_1= parent_filed_2 cost_filed_1= cost_filed_2; *parent_filed_2*=null *cost_filed_2*=null

Step 8: end of while loop

Step 9: reiteration till when extent at the sink node

V. RESULTS AND DISCUSSION

After In this section we present experiment analysis using log file, Once simulation has finished it will create trace file at background, which contains all information of node communication as well as other log information. we have created a database of 5 text files which contains reading of 5 ms each till 25 ms as our simulation time is 25 ms. After that, we read the text file in a program created for the trace in Netbeans IDE 8.2. We got readings of various events in Netbeans from which we have plotted the graph of various parameters such as Drop rate(DR),Throughput and Packet Delivery Ratio(PDR) calculate d according to equation (1), (2) and (3) respectively. The evaluation has done with various WSN as well as cluster network existing protocol [12] [13] [14].

Table 1. The simulation parameters has used which is described in below table

Parameter	**Values**
Simulator	NS-allinone 2.35
Simulation time	25 sec
Channel Type	Wireless Channel
Propagation Model	Two Ray Ground
Standard	MAC/802.11

Simulation Size	1000 *1500
Max packet Length	1000
Ad hoc routing	AODV
Traffic	CBR

Table 2: shows the basic difference between the proposed and existing WSN.

Parameters	**WSN [21]**	**Proposed (cluster base with AODV)**
Data Aggregation	No	Yes
Data Security	Yes (selective)	Yes
Energy Conservation	No	Yes
Packet Loss	High	Low
End to End delay	High	Low
Packet Overhead	High	Low

1. Drop Rate:

It is defined as the number of packet lost per number of packet sent. The smallest
amount value of drop rate states superior performance of the protocol.

$$Drop\ Rate = \sum_{i=0}^{n} \left(\frac{packet\ received\ [i \ldots n]}{sent\ packet\ [i \ldots n]} \right) \quad \ldots\ldots (1)$$

2. Throughput:
It is defined as the total number of packets supplied over the entire simulation era.
This is a mix of TCP's total number of packets and the total number of packets sent. The higher throughput value means higher performance of the protocol.

$$Throughput = \left(\frac{\sum_{i=0}^{k} received\ packet[TCP]}{\sum_{i=0}^{l} sent\ packet[TCP]} \right) * 100 \quad \ldots\ldots (2)$$

3. Packet Delivery Ratio (PDR):
The PDR is defined as the ratio of the data packet numbers to the number of packets
produced through the network. The higher value of the packet distribution ratios represents greater efficiency of the protocol.

$$PDR = \sum_{i=0}^{n} \left(\frac{packet\ received\ [TCP]}{packet\ sent\ [TCP]} \right) * 100 \quad \ldots\ldots (3)$$

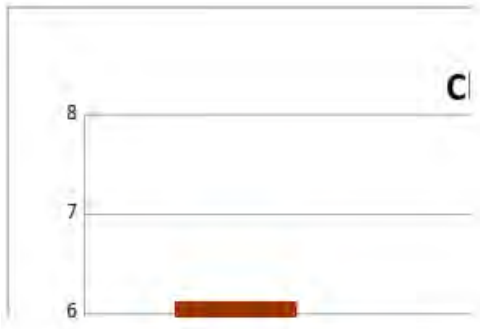


Figure 3: Drop rate of proposed vs existing
This figure 3 will provide the drop rate overall simulation during the communication with other protocols. The above graph has calculated based on various experiment analyses in NS2 environment. All defined protocols has used with different number of nodes in cluster network. The graph shows minimum packet drop rates of proposed AODV than other protocols.

Figure 4: Throughput of proposed vs existing
This figure 4 will provide the throughput of system during the communication with other protocols. Throughput s the most vital parameter the measure the QoS of any network. Similar as first experiment .tr files has utilized for evaluate the throughput
for all protocols. This figure also shows our approach produces highest throughput
than all three protocols.

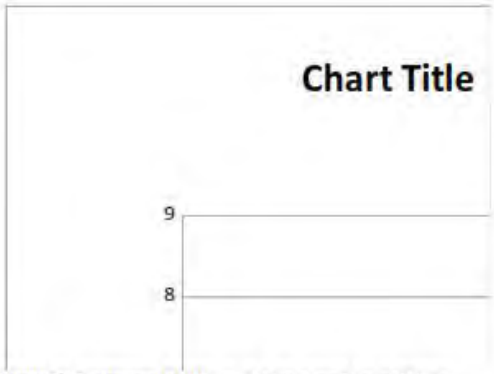


Figure 5: Network lifetime of proposed vs existing
This figure 5 will provide the how system will increased the actual time percentage of simulation due to proposed energy conservation protocol. Network lifetime has calculated based on after applying the energy conservation protocol, conservation s the technique during the data transmission eliminate unnecessary energy utilization by sensor, receiver as well as internal nodes respectively. In proposed system BTC has
used for best path selection as well as energy conservation protocol.

VI. CONCLUSION

This proposes an aggregation of the HMAC and data in WSN. The CHs are
initially chosen based on the connectivity of the node, which acts as a Data
Aggregator. The proposed techniques provide the ad hoc network with greater security and also prevent different types of network attacks. It supports in the direction of
escalations the packet delivery ratio (PDR) as well as decreases the network directly above by cultivating the enactment of the respective routing protocol. In the future, for faster detection of wormhole nodes, it will also need to modify the table entries in the receiver node. And also improved the security of the wireless ad hoc network. By means of set up such effective approach to prevent various kin fog network attacks with new defined algorithm. It also exposure and

elimination of wormhole attack throughout data communication. The proposed Techniques offer greater security for various types of ad hoc networks and also deter network assaults. This facilitates the elevation of the packet distribution ratio (PDR) and reduces the overhead network by encouraging the enactment of the respective routing protocol. To get the detection of wormhole nodes faster, the table entries at the receiver node need to be improved in future plan as well. And the security of ad hoc wireless networks is also improved. By applying such an effective approach, the kin fog network will deter various attacks with new defined algorithm.

References

[1] Yashpalsinh Gohil, Sumegha Sakhreliya, Sumitra Menaria "A Review On: Detection and Prevention of Wormhole Attacks in MANET", International Journal of Scientific and Research Publications, Volume 3, Issue 2, February 2013

[2] Vikaskumar Upadhyay,Rajesh Shukla "An Assessment of Wormwormhole attack over Mobile Ad-Hoc Network as serious threats", Int. J. Advanced Networking and Applications Volume: 05 Issue: 01,2013

[3] Dhruvi Sharma, Vimal Kumar, Rakesh Kumar, " Prevention of Wormhole Attack Using Identity Based Signature Scheme in MANET " , Computational Intelligence in Data Mining. Volume 2. Volume 411 of the series Advances in Intelligent Systems and Computing pp 475-485. 10 December 2015, Springer.

[4] Elham Zamani and Mohammadreza Soltanaghaei, " The Improved Overhearing Backup AODV Protocol in MANET" , Journal of Computer Networks and Communications Volume 2016 (2016), Article ID 6463157, 8 pages. Hindawi

[5] Amit Kumar, Sayar Singh Shekhawat, " A Parameter Estimation Based Model for Wormwormhole Preventive Route Optimization" , IJCSMC, Vol. 4, Issue. 8, August 2015, pg.80 – 85 International Journal of Computer Science and Mobile Computing.

[6] https://en.wikipedia.org/wiki/Wireless_sensor_network

[7] Ashish Kumar Jain, Ravindra Verma, " Trust-Based Solution For Wormhole Attacks In Mobile Ad Hoc Networks" , Volume-4, Issue-12, November- 2015, Global Journal Of Multidisciplinary Studies

[8] Rajan Patel, Anal Patel, Nimisha Patel, " Defending Against Wormhole Attack in MANET" , Fifth International Conference on Communication Systems and Network Technologies, 2015 IEEE.

[9] Dhruva Patel, Parth Trivedi, M. B. Potdar, " A brief Analysis on Detection and Avoidance Techniques of Wormhole Attack in MANET" , Volume 117 - Number 16, International Journal of Computer Applications, 2015.

[10] Juhi Biswas, Ajay Gupta, Dayashankar Singh, " WADP: A wormhole attack detection and prevention technique in MANET using modified AODV routing protocol", 9th International Conference on Industrial and Information Systems (ICIIS), 2014. IEEE.

[11] Kapil Raghuwanshi, Amit Saxena, Manish Manoria, " An Enhanced Integrated Solution for Identification and Elimination of Wormhole Attack in MANET" ,Volume 110 - Number 7, International Journal of Computer Applications 2015.

[12] Uddin MA, Islam MM, Khanom R, Mosaddek M. Efficient and Secure Cluster based Routing Protocol for Wireless Sensor Network.

[13] Ouafaa I, Mustapha E, Salah-Ddine K, SAID EH. PERFORMANCE ANALYSIS OF SLEACH, LEACH AND DSDV PROTOCOLS FOR WIRELESS SENSOR NETWORKS (WSN). Journal of Theoretical & Applied Information Technology. 2016 Dec 31;94(2).

[14] Upadhyay R, Bhatt UR, Tripathi H. DDOS attack aware DSR routing protocol in WSN. Procedia Computer Science. 2016 Mar 1;78(C):68-74.

[15] Hu, F.; Hao, Q.; Bao, K. A Survey on Software-Defined Network and
OpenFlow: From Concept to Implementation. IEEE Commun. Surv. Tutor. 2014, 16,
2181–2206.

[16] Gante, A.D.; Aslan, M.; Matrawy, A. Smart wireless sensor network management based on software-defined networking. In Proceedings of the 2014 27th Biennial Symposium on Communications (QBSC), Kingston, ON, Canada, 1–4 June
2014; pp. 71–75.

[17] Olivier, F.; Carlos, G.; Florent, N. SDN Based Architecture for Clustered WSN. In Proceedings of the 2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), Blumenau, Brazil, 8–10 July
2015; pp. 342–347.

[18] De Oliveira, B.T.; Margi, C.B.; Gabriel, L.B. TinySDN: Enabling multiple controllers for software-defined wireless sensor networks. IEEE Lat. Am. Trans. 2015,
13, 3690–3696.

[19] Tootoonchian, A.; Ghobadi, M.; Ganjali, Y. OpenTM: Traffic matrix estimator for OpenFlow networks. In Proceedings of the International Conference on Passive and Active Network Measurement, Zurich, Switzerland, 7–9 April 2010; Springer: Berlin, Germany, 2010; pp. 201–210.

[20] Jain, R.; Paul, S. Network virtualization and software defined networking for cloud computing: A survey. IEEE Commun. Mag. 2013, 51, 24–31.

[21] Zhou, J.; Jiang, H.;Wu, J.;Wu, L.; Zhu, C.; Li,W. SDN-Based Application
Framework forWireless Sensor and Actor Networks. IEEE Access 2016, 4, 1583–1594

Hybrid Cryptography Approach For Secure Sharing Of Data using AES, ECC & SHA256

Pooja Patil, Dr. Rajesh Bansode
Information Technology, Thakur College of Engineering & Technology
ppatil879@gmail.com, rajesh.bansode@thakureducation.org

Abstract— **Nowadays, the networks have gone global and information has taken the digital form. Critical information now gets stored, processed and transmitted in digital form on computer systems and open communication channels. Since information plays such a vital role, hackers are targeting the computer systems and open communication channels to disrupt the critical information system. Different cryptographic techniques like symmetric and asymmetric algorithms are used to hide any confidential information from the attacker. Symmetric key algorithms are simple and faster, but easily cracked by attacker. Asymmetric key algorithms have higher security, but lack in time efficiency. But, when symmetric and asymmetric algorithms combines together, they provides better result, and this technique known as hybrid cryptography. Using the concept of hybrid cryptography technique, we proposed hybrid system which combines symmetric algorithm (AES) and asymmetric algorithm (ECC) together, that can provide greater security and less execution time for encrypting and decrypting plaintext. Also cryptographic hash function (SHA256) implemented to evaluate the integrity of data, authentication control and other security mechanisms.**

Keywords— Cryptography, AES, ECC, SHA256, Encryption, Decryption.

I. INTRODUCTION

The spectacular growth of the Internet has make an increased awareness of an interest in security issues. Security of data is primary aspect when it transfer over the internet and protecting over data from unauthorized access is very necessary. To keep information secret, a widely used method is an encryption and decryption. Basically, encryption/decryption are the fundamental functions of cryptography In encryption, a simple message (plain text) is converted into unreadable form called ciphertext. While in decryption, a ciphertext is converted into the original text (plaintext)[1].

Different Cryptographic techniques were implemented for securing data during various communications. Cryptography algorithms mainly divided into two categories such as Symmetric key cryptography and Asymmetric key cryptography. In the symmetric key encryption, same key is used for both encryption and decryption process. Symmetric key algorithms are AES, DES, 3DES, Blowfish etc [1],[2]. Asymmetric key encryption is the technique, in which the different keys are for the encryption and the decryption process. One key is public and second is kept private. Asymmetric key algorithms are RSA, ECC, ElGamal, DSA etc [1],[2]. And the concept which combines both symmetric and asymmetric cryptographic techniques is known as hybrid cryptography. A hybrid cryptosystem is a protocol using symmetric and asymmetric cryptographic technics together, each to it's best advantage.

The main objective of cryptographic algorithms is that to transmit the data speedily and in a secure manner with confidentiality, authentication and integrity [2]. The another objective of the research is to provide encrypt & decrypt data efficiently and effectively protect the transmitted data along with improved performance parameters.

Also, Hash functions are important cryptography building block. They combines strengths of both methods (symmetric and asymmetric-key cryptosystems) [3]. Hash functions are used to assure integrity and authentication. This system will provide better security to data transfer over the internet.

II. LITERATURE SURVEY

Survey of various research papers done in this section. Different cryptographic algorithms and Hash functions are analyzed with different parameters by various researchers.

In paper [1], different types of symmetric and asymmetric algorithms are explained in detail. Also, Different parameters analyzed such as key generation time, encryption and decryption time etc. Performance evaluation can be improved in future.

In paper [2], the main objectives of cryptography such as confidentiality, integrity, authentication are explained. Comparison of different types of algorithm is done on the basis of performance parameters and it is concluded that ECC provides better security with high speed.

In paper [3], SHA and MD5 algorithms are explained in detail. It is concluded that SHA1 is more faster than MD5, and SHA2 is even more secure than SHA1 and MD5. Hash functions can be implemented with hybrid cryptography to improve performance & security in future.

In paper [4], encryption and decryption of text using AES is explained. It provides security to multiple accounts, multiple files which have confidential data. Security can be improved in future.

In paper [5], encryption and decryption of text using ECC is explained. Encryption and decryption time analyzed for text. It is concluded that ECC has low power consumption, less memory requirement, small key size and high security.

In paper [6], Image encryption and decryption using AES algorithm is implemented to secure the image data from an unauthorized access. The original images can also be completely reconstructed without any distortion. It is concluded that it has extremely large security key space & can withstand with attacks like the brute force attack, cipher attacks and plaintext attacks.

In paper [7], Image encryption and decryption using ECC algorithm is implemented to provide authenticity and integrity. Also, speed analysis & correlation analysis discussed.

In paper [8], hybrid cryptography approach implemented to provide encryption to the multimedia data
such as text, image, audio, video which resulted in an
output with 100 percent accuracy without any loss of information.

III. PROBLEM DEFINITION

Cryptographic techniques provides the secure data transmission, but there are some complexities in existing systems. Most of the cryptographic techniques are time consuming processes. Some techniques does not includes the integrity checks on transmitted data. The amount of cipher text is higher than the original text. Another issue is lack of security during key exchange.
In order to implement an effective cryptographic algorithm all these aspects has to be considered in order to make it robust. There is need to implement the technique which helps to overcome such complexities in such existing system.

The proposed hybrid cryptographic technique uses the best features of symmetric (AES) and asymmetric (ECC) cryptographic technique with hash function (SHA256). So that, this technique will help to reduce space and time complexity. It will provide the validation of data integrity. Also, data transmission can be more efficient, if cipher size is reduced.

IV. DISCRYPTION OF AES, ECC & SHA2

A. Advanced Encryption Standard(AES)

The Advanced Encryption Standard is a block cipher which takes as input a 128 bit plaintext, which is subject to an encryption with 128 , 192 and 256 bit key depending upon the number of rounds i.e 10,12 or 14 respectively [9].
Fig.1 explains the various internal rounds that take place for encryption and decryption using AES. It broadly consists of Substitution i.e bit by bit substitution, shifting of rows i.e transposition, mixing of columns based on modular arithmetic multiplication followed by adding of round key till n-1 rounds. Mix column round is omitted in
the final nth round. After the nth round a 128 bit cipher text is obtained.

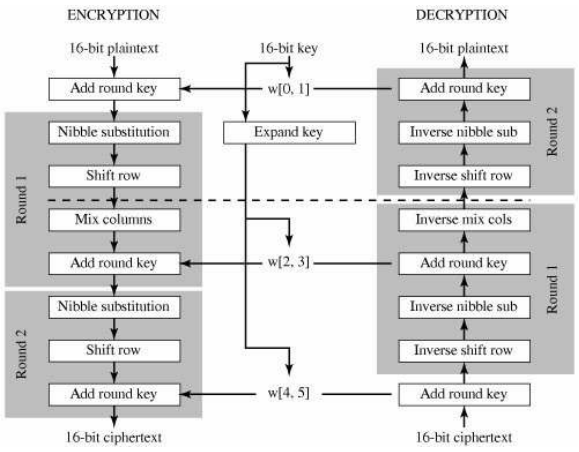


Fig.1 Block diagram of Advanced Encryption Standard (AES) for encryption & decryption

B. Elliptic Curve Cryptography(ECC)

ECC provides better security with a smaller key size if we compare it with other asymmetric algorithms.. High level of security can be achieved using a small key size. ECC works on elliptic curve equation [10]. Elliptic curve equation for binary field is written as- $y^2+xy=x^3+ax+b$ Where a and b are two constants, different elliptic curves will be shaped with different values of these two constants. Elliptic curve equation for prime field is given as- $y^2=x^3+ax+b \bmod p$

Here a and b are constants and p is a prime number. Grater the value of number p more will be the number of points generated on the elliptic curve. Large number of points on the curve gives high level of security. Elliptic curve is shown in figure below-

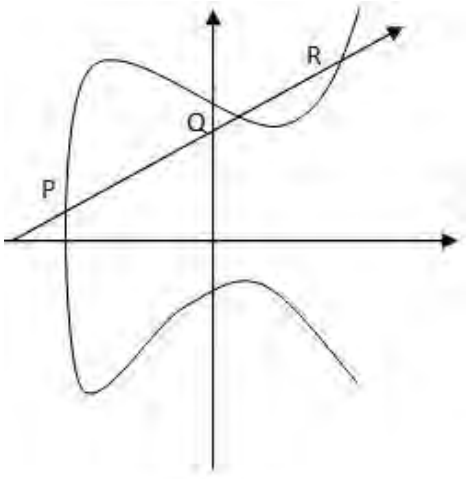


Fig.2 The algebraic structure of elliptic curve on which Elliptic Curve
Cryptography (ECC) is based on

C. Secure Hash Algorithm 256 (SHA256)

SHA-2 is a set of cryptographic hash functions the variants of SHA-2 are SHA-224, SHA-256, SHA-384, SHA-512, SHA-512/224, SHA-512/256. SHA-2 includes a significant number of changes from its predecessor,

SHA-1. SHA-2 currently consists of a set of six hash functions with digests that are 224, 256, 384 or 512 bits [11].

V. PROPOSED METHODOLOGY

The proposed system consists of two processes such as encryption process and decryption process. It uses symmetric algorithm AES and asymmetric algorithm ECC along with hash function SHA256. They are explained as follow:

ECC a public-key algorithm can be used for encryption as well in digital signature processes. The key management is an essential feature of ECC algorithm. ECC provides better security with small key size [12]. The security of the method used in ECC is based on the difficulty of factoring large numbers.

AES is not only a secure cipher but it offers a very high performance and makes better use of resources. Operational cost is less and less memory requirement. It is strong enough to be certified for use by the US government for top secret information Encryption Process. Not a single successful brute-force attack on AES has been found till date, the only possible known attack against AES.

SHA-256 a message digest function with a block size of 512- bit generates 256-bit message digest. All major SSL certificate issuers now use SHA256 which is more secure and trustworthy than SHA1. SHA256 is currently much more resistant to collision attacks as it is able to generate a longer hash which is harder to break.

So all the essential features of these algorithms are made available in our hybrid algorithm [13]. Better encryption of AES, most efficient key management by ECC along with the digital signature by making use of SHA-256 are included in a single hybrid system.

A. Encryption Process

1. An AES key 'K' of 128-bit, 192-bit or 256-bit is chosen.
2. Encrypt message (M) using AES algorithm and above selected key K.

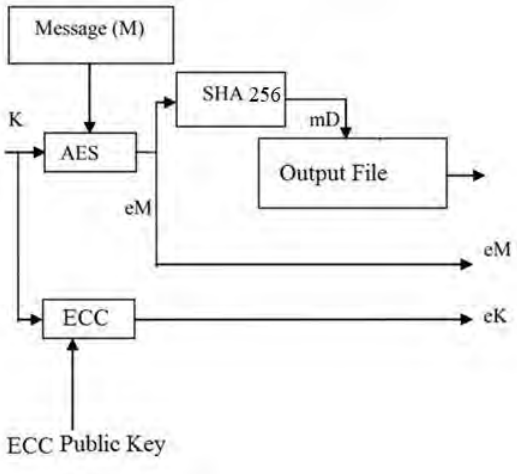


Fig. 3 Encryption process of proposed system

B. Decryption Process

This process is the reverse of encryption process and is having following steps:

1. The encrypted AES key (eK) is decrypted with ECC algorithm.

K = ECC-decryption (eK)

2. Similarly the encrypted message (eM) is decrypted by

AES algorithm using key K

M = AES-decryption (eM)

3. The message digest of encrypted message (eM) is computed using SHA256.

mD = SHA256 (eM)

4. Thus we get message (M) of sender eM = AES-encryption(M)

3. AES key K is encrypted by making use of ECC algorithm.eK = ECC-encryption (K)

4. The cipher text (eM) is fed to SHA256 algorithm which generates a message digest of 256-bit.

mD = SHA256 (eM)

5. The encrypted message (eM) and AES encrypted key (eK) is transmitted to the user over a network.

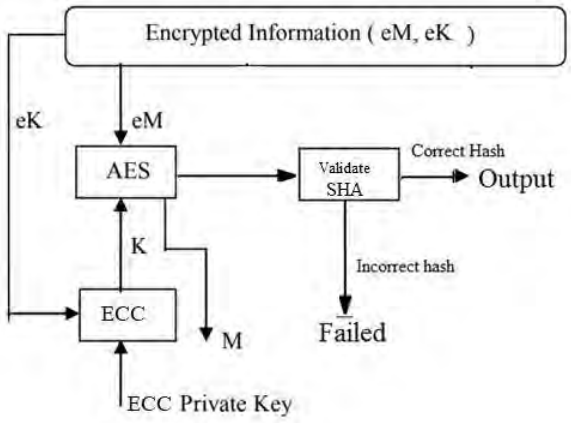


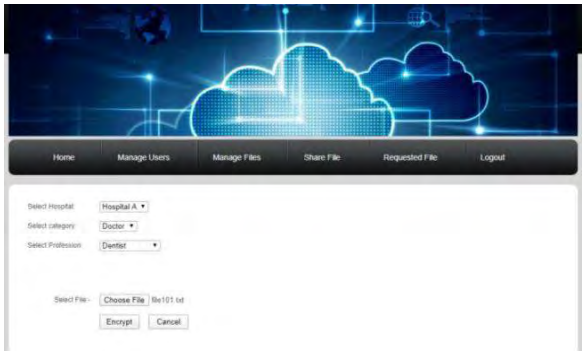
Fig. 4 Decryption process of proposed system

Proposed hybrid cryptography method is implemented for secure sharing of patient healthcare information on website. Encryption and decryption time taken for different text and image files are calculated. Also, hash matching is performed.

There are following steps:**Step 1:** Admin login into website using login credentials such as username and password

Step2: After successful login, multiple users added by the admin. So the admin can share files with user.

Step3: File is selected by the admin for uploading purpose.



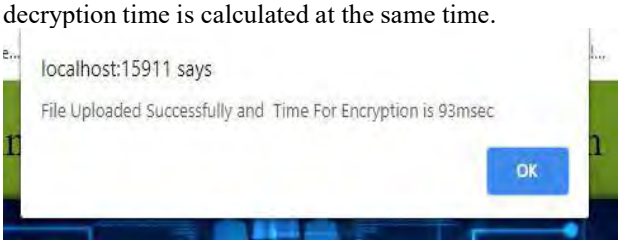
Step4: File is uploaded on website and encryption time is calculated.

Step5: File is successfully shared with user by the admin.

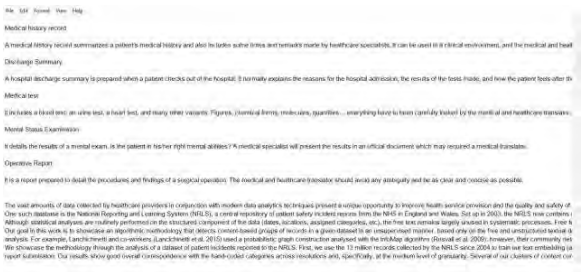
Step6: User login into website using login credentials such as username and password.

Step7: User can view shared files which are sent by the admin.

Step8: User enters the secret key to decrypt the file and decryption time is calculated at the same time.



Step9: After successful decryption, file can be downloaded.



Step10: To perform hash matching, hash is generated to check whether it is matching with its hash value or not.

VI. RESULT AND ANALYSIS

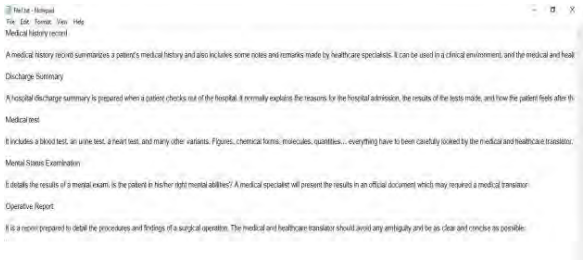
This methodology implemented for secure sharing of healthcare data and can be used for another web applications also. Different types of text (txt, docx, pdf) and image (jpg, png) files are taken as input. Encryption is performed on original file and it sent to the intended receiver. Receiver decrypted that file using secret key and matching of hash is performed.

Text and Image files:

1. Text File:

(a) Original text file

(b) Encrypted text file



(c) Decrypted text file

2. Image File:

(a) Original Image

(b) Encrypted Image

(c) Decrypted Image Encryption and Decryption time of different types of files are calculated and mentioned in following table:

TABLE I
Encryption And Decryption Time

File Type	File Size	Encryption Time (ms)	Decryption Time (ms)
.txt	4KB	50	107
.docx	24KB	57	86
.pdf	1880KB	390	142
.jpg	296KB	92	91
.png	7KB	62	92

VII. CONCLUSION

The limitations of various cryptographic techniques are analyzed and we proposed this hybrid model which is the combination of the AES, ECC and SHA256. This system performs encryption and decryption for better security of confidential data. It provides enhanced time and validation of data integrity.

References

[1] Faiqa Maqsood,Muhammad Ahmed, Muhammad Mumtaz Ali,Munam Ali Shah, ‐Cryptography: A Comparative Analysis for Modern Techniques‖,‐International Journal of Advanced Computer Science and Application‖, Vol: 8, Issue: 6, 2017

[2] Hemangi Zope, Prof. Savita Sangam, ‐Comparative Analysis of Various Encryption Algorithms and Techniques‖ International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; Volume 5, Issue XII, December 2017

[3] Surbhi Aggarwal ,Neha Goyal,Kirti Aggarwal, ‐A review of Comparative Study of MD5 and SHA Security Algorithm‖, International Journal of Computer Applications (0975 – 8887) Volume 104 – No.14, October 2014

[4] Kundan Kumar Saraf, Nandini Goyal, Shubham Raokhande, Akash Takale,‖ Text Encryption and Decryption using AES Algorithm‖, IJEECS, ISSN 2348-117X, Volume 7, Issue 3 March 2018

[5] Keerthi K and Dr.B.Surendiran, ‐Elliptic Curve Cryptography for Secured Text Encryption‖, In International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2017

[6] Priya Deshmukh, ―An Image Encryption and Decryption Using AES Algorithm‖, International Journal of Scientific & Engineering Research(IJSER), Vol.7, Issue.2, pp.210-213, 2016.

[7] Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh, ‐Image encryption using elliptic curve cryptography‖, Eleventh International Multi-Conference on Information Processing-2015, Procedia Computer Science, vol. 54, pp. 472-481, 2015.

[8] Sridhar C.Iyer, R.R.Sedamkar, Shiwani Gupta, ‐An Efficient Multimedia Encryption using Hybrid Crypto Approaches ,‖ International Journal Of Recent Trends in Engineering & Research, 2016 , pp. 442-
452, Volume 02.

[9] Sridhar C. Iyer, R.R. Sedamkar, Shiwani Gupta, ‐A Novel Idea on Multimedia Encryption Using Hybrid Crypto Approach,‖ 7th International Conference on Communication, Computing and Virtualization-2016 (ICCCV-2016), Procedia Computer Science; Vol 79, p.293-298, ISSN: 1877-0509

[10] Samiksha Sharma, Vinay Chopra, ‐Analysis of AES Encryption with ECC,‖ Proceedings of International Interdisciplinary Conference on Engineering Science & Management, 17th - 18th December 2016, pp.
195-201.

[11] Mahesh A. Kale, Prof. Shrikant Dhamdhere, ‐Survey Paper on Different Type of Hashing Algorithm‖ ,IJASRET, Volume 3 ,Issue 2,| February 2018 , ISSN (Online) 2456-0774

[12] Soumi Banerjee and Anita Patil ,‐ECC Based Encryption
Algorithm for Lightweight Cryptography‖,2018, Springer

[13] Divija Ameta, Sandeep Upadhyay,‖A Hybrid Approach for Image Encryption using Different Number Iterations in ECC and AES Techniques‖, International Journal of Compter Applications (0975 –
8887) Volume 175 – No.3, October 2017

Securing IoT Devices Generated Data using Homomorphic Encryption

Mrs. Anita Caudhari, Dr. Rajesh Bansode
Mumbai University
rajeshanitac@sjcem.edu.in, bansode1977@gmail.com

Abstract –.**Cloud computing is emerging trends Cloud service provider provides services like,network,application,storage etc.Peoples use Clod storage for stor- ing and retrieving their data for future use. Internet loT devices have memory and power constrained .So, all data generated by these devices are outsourced on cloud .But sharing private data to cloud is risky. Though it is in encrypted format. Processing data on cloud is import task .Our Research focuses on processing on Encrypted data .we had developed GSW scheme to perform computation an en- crypted data.Our scheme takes data from IoT devices and encrypt it using GSW algorithm. Computation is performed on encrypted data like addition and multi- plication.**
Keywords: HE,Cloud,Operation

I. INTRODUCTION

Homomorphic encryption method carried out on encrypted data that is cipher text instead of plain text .The main thing in homomorphic encryption is after performing computation on encrypted data ,generated result should match with the computation performed on plain-text data. In following figure, we are performing simple design of homomorphic addition where plain-text is 4 and 7 and generated result 11 when we do encryption of 4 and 7 by using simple method just multiply plaintext by 3 we will get encrypted result 1 and 21. at decryption side we will use method cipher-text divide by

3 we will get original result. IoT devices delegate encrypted data to cloud ,If some computation are required on data, user has to download data and perform computation. This problem can be solved by using homomorphic encryption on cloud side[1,2]. For Preserving privacy of data when sending data to cloud has to encrypt plaintext before sending data to cloud. This might bound the value of the information , but current trends in cryptography permit probing on encrypted information and execution of different actions on encrypted data, without revealing secret information [3]. An encryption system which permits random actions on ciphertexts known as a Fully Homomorphic Encryption (FHE) scheme. The initial FHE system was constructed by Gentry, and several succeeding systems have quickly developed more practical, with enhanced performance and parameters. Subsequent FHE systems have called Somewhat Homomorphic Encryption (SHE) system as an fundamental structure and use a method called bootstrapping to spread in FHE scheme[4]. computation on encrypted data is ongoing research different authors propose different algorithm to reduce computational cost. This type of model can be used in healthcare application to secure patient data[5].

II. LITERATURE SURVEY

Public key encryption with keyword exploration based homomorphic encryption in multi user situation. In this method DGHV homomorphic encryption applied cloud server generate reversed encryption directory structured. DHGV not uses query trap door. Permits several users to achieve encrypted keyword search over encrypted data. If number of queried keyword is 300 takes 9 s[6].The encoded information can be considered as a good solution to forget over these impediments on computing. Researchers have proposed another encryption form, homomorphic encryption (HE), that the third-party have the capacity to complete tasks on encrypted information. The homomorphic property makes HE schemes helpful in a extensive scope of protection saving applications by the homomorphic property, for example, electronic voting and advanced human services. The Homomorphic encryption method was presented in 1978.

From that, numerous HE schemes have been proposed. HE plans can be isolated into two classes: somewhat and fully homomorphic encryption. The first categories give one homomorphic property[7]. A few homomorphism properties are offered by the full

homomorphic. Since re-appropriated scrambled information under homomorphic schemes can be put away for a long-lasting, more often not by the equivalent key, in CC conditions and since the cloud customers frequently access to cloud services utilizing asset restricted gadgets. These plans should be advanced as far as security level and running time to work successfully. Designed FPGA homomorphic encryption co- processor [8].

Cloud computing is used to convey on-request facilities (e.g., capacity, applications, networks, servers) using the Internet. This is because of a few benefits given by cloud providers containing low costs, high administrations, power calculations flexibility and adaptability. Regardless, failures over the privacy of sensitive information are as yet the main drawback constraining the adoption of Cloud computing services[9]. Translating algorithm that can run on encrypted data. translate basic operators like bitwise, arithmetic and relational operators. Also discussed decision making, loop handling data structures like stack, queue, in that how these algorithm can be applied on encrypted data. Processed each algorithm and measured time required to process encrypted data[10,11].

Hybrid homomorphic encryption model that combines public key encryption and somewhat homomorphic encryption for reducing storage requirement. Here message are encrypted using PKE and computation carried out using FHE or SHE. Uses public key size 3TB,multiplicative depth 20,comparative analysis of proposed algorithm with SHE only and AES with SHE[12]. Homomorphic proxy re-encryption scheme permits different users to share data they outsourced homomorphically encrypted using separate public key. Uses Damgard -Jurik cryptosystem. Evaluated the performance of proposed model in terms of time computation and encrypted data volume. Delegator will take

0.004 sec ,proxy ,i.e cloud 120sec ,delegator 30 sec for 92x122 pixel[13].

Secret keys implemented a lattice-based Key-Policy Attribute-Based Encryption (KP- ABE) scheme. The specific KP-ABE plan can be utilized straightly for Attribute-Based Access Control (ABAC) applications. Graphics Processing Units (GPUs) is the state- of-the-art implementation and demonstrates that the homomorphic open key and ciphertext evaluation activities, which command the execution time of the KP-ABE plot, can be performed in a sensibly short measure of time [14].

Privacy-Preserving Data Processing (PPDP) framework with the help of a Homomorphic Re-Encryption Scheme (HRES). The HRES expands fractional HE from a solitary client framework to a multi-client one by offering ciphertext re-encryption to enable numerous users to get handled ciphertexts. PPDP system can maintain seven essential activities over ciphertexts, which incorporate Addition, Subtraction, Multiplication, Sign Acquisition, Comparison, Equivalent Test, and Variance [15]. Implemented optimized RLWE scheme in Bayesian spam filter, Decision tree, secure multiple keyword search.

Alhassan Khadr et,al(2017) developed model for securing healthcare data

.maintaining patient data privacy is important. Focused on NTRU based implementa- tion on GSW HE and results shows improvement of 58x in CPU performance. Per- formed Operation on blood pressure data, person goodness of fit test[16]-health.to se- cure patient data proposed secure privacy preserving data aggregation scheme. Based on bilinear pairing .Here focused on improving data aggregation efficiency and data privacy. Reduced communication and computation overhead.

Described that an integrated individual health data framework that permits secure capacity and prepared the medical information in the cloud by utilizing an exhaustive homomorphic encryption model to protect information security[20,21].Medical Cyber Physical system based on different algorithm proposed to secure patient data. ECIECS,KP-ABE,CP-ABE,PAILIER ,FHE algorithm used in securing healthcare data but main limitation is computation time[17].Provided solution to protect users data di-rectly into the cloud while preserving the user privacy

.uses cardiac risk factor algo- rithm on encrypted data.to execute this algorithm it take 3.2 second.FHE is imple- mented on healthcare data .perform computation on encrypted heart rate, provided av- erage heart rate ,LQTS,Min,Max Heart rate[18].

III. METHODOLOGY

Homomorphism If (F1, ∗) and (F2, ⊗) are two groups, then a function f : F1 → F2 is a group homomorphism if f(a ∗ b) = f(a) ⊗ f(b) for all x, y ∈ F1 Examples: f(a) = e a , f(a) = log(a),...

GSW:

A ciphertext C encrypts a message m if the following holds:

$$C\text{iphertext} * vco = m\text{message} * vco + err \quad (1)$$

where err is a small error term
Can decrypt if vco has a "big" coefficient vi by rounding

$$A=\{< Ciphertext\ i, , vco >\} \quad (2)$$

$$B=\{< message * vcoi + err >\} \quad (3)$$

$$A \div Vcoi = B/Vcoi \quad (4)$$

where Ciphertext i denotes the i th row of Ciphertext

Ciphertext 1 encrypts message1 and ciphertext 2 encrypts message2,

3.1 Addition:

Ciphertext1 + Ciphertext2 encrypts $m$essage1 + $m$essage2 (5)
Ciphertext1 + Ciphertext 2 vco = $m$essage1 + $m$essage2 vco + err1 + err2 (6)

3.2 GSW Key Generation

Secret Key: $t \xleftarrow{\$} Z_q^n$

Public Key A:

$$B \xleftarrow{\$} Z_q^{m*n}$$
$$err \xleftarrow{\$} X^m$$

A*s = Bt + err− Bt =err (10)

Encryption:

For plaintext message € Zq

Ciphertext<- F($m$essage · IN + BitDecomp (R ·A)) (11)

IV. EXPERIMENTAL RESULTS

A Homomorphic Encryption is the transformation of data into ciphertext that can be examined and functioned as if it were still in its unique form. Homomorphic encryption shows a significant part in cloud computing, permitting patients to store encoded PHR files in a public cloud and take advantage of the cloud provider's analytic services. The scheme prevents rogue insiders from violating privacy and avoids unintentional leakage of private data. Homomorphic Encryption schemes are utilized to achieve processes on encoded information without knowing the private key (i.e., without decryption), the client is the only holder of the secret key. Once the result of the process is decrypted, it is the equivalent as if it had carried out the calculation on the raw data.

In this section we will discuss about results of homomorphic encryption on data set. Collected from IoT devices and performed GSW encryption on it .

Table 1.Time Taken for encryption and Decryption based on Record size(in Second)

Record Size	Encryption Time(First Field)	Encryption Time(Second Field)	Decryption Time(Addition)
50	2.40	2.43	2.46
100	111.97	111.98	112.00
200	233.42	233.43	233.45
300	382.32	382.33	382.37
500	648.47	648.49	648.50

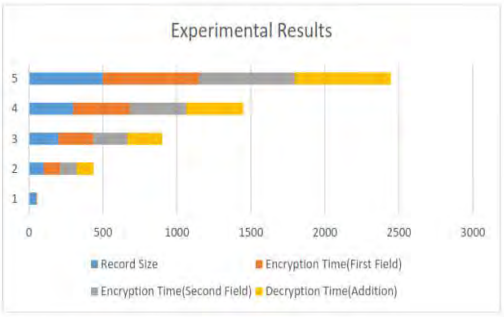


Figure 1. computation time based on record size
If we increase security parameter, there will be exponentially increase in time for performing operation.

V. Conclusion

In this research, we had collected data from IoT devices and after encryption uploaded it on to cloud server. User sends query to cloud for performing operation on encrypted data. Computation will be performed once query will be fired on cloud. Here we are using only addition and multiplication operation .Computation time varies based on data set and

security parameters. If noise will be more then ,not getting proper result. In future focus will be on comparative operators.

References

[1] Abbas A.,Hidayet Asku.,"A Survey on Homomorphic Encryption Schemes :Theory and Implementation ",*ACM Computing Survey* ,Vol. 51 ,No. 4,Article 79,July 2018.

[2] Çetin, G. S., Chen, H., Laine, K., Lauter, K., Rindal, P., & Xia, Y. ,"Private queries on Encrypted Genomic Data*", BMC Medical Genomics*", 10(2), 45,2017.

[3] Wang, B., Zhan, Y., & Zhang, Z. , "Cryptanalysis of a Symmetric Fully Homomorphic Encryption Scheme", *IEEE Transactions on Information Forensics and Security*, 13(6), 1460-1467,2018.

[4] Cousins, D. B., Rohloff, K., & Sumorok, D., "Designing an FPGA-accelerated homo- morphic encryption co-processor", *IEEE Transactions on Emerging Topics in Compu- ting,* 5(2), 193-206,2017.

[5] Wu, D. N., Gan, Q. Q., & Wang, X. M. ," Verifiable Public Key Encryption With Keyword Search Based on Homomorphic Encryption in Multi-User Setting", *IEEE Access* , 6, 42445-42453,2018.

[6] Wang, B., Zhan, Y., & Zhang, Z. (2018). Cryptanalysis of a Symmetric Fully Homo- morphic Encryption Scheme. IEEE Transactions on Information Forensics and Secu- rity, 13(6), 1460-1467.

[7] Ayantika C,Indranil S,"Translating Algorithm to handle Fully Homomorphic En-crypted data on Cloud*",IEEE Transaction on Cloud Computing*,2015

[8] Jung Hee C ,Jinsu K,"A Hybrid Scheme of Public Key Encryption & Some what Ho- momorphic Encryption",*IEEE Transaction on Information Forensics and Secu- rity*,2015

[9] Reda B,Gouenou C,"Proxy Re-Encryption Based on Homomorphic Encryption",*ACM* ,2017

[10] Ding, W., Yan, Z., & Deng, R. H. ," Encrypted data processing with homomorphic re- encryption", *Information Sciences*, 409, 35-55,2017.

[11] Alhassan K,Glenn Gulak,"SHIELD:Scalable Homomorphic Implementation for En- crypted Data Classifiers",*IEEE Transactions on Computers* ,2015 .

[12] Alhassan K, ,Glenn Gulak,"SecureMed:Secure Medical Computation using GPU- Accelerated Homomorphic Encryption Scheme*",IEEE journal of Biomedical and Health Informatics*,2017

[13] Ara, A., Al-Rodhaan, M., Tian, Y., & Al-Dhelaan, A.," A Secure Privacy-Preserving Data Aggregation Scheme Based on Bilinear Elgamal Cryptosystem for Remote Health Monitoring Systems", *IEEE Access*, 5, 12601-12617,2017.

[14] Dai, W., Doröz, Y., Polyakov, Y., Rohloff, K., Sajjadpour, H., Savaş, E., & Sunar, B., "Implementation and Evaluation of a Lattice-Based Key-Policy ABE Scheme", *IEEE Transactions on Information Forensics and Security*, 13(5), 1169-1184,2018.

[15] Bocu, R., & Costache, C.," A Homomorphic Encryption-Based System For Securely Managing Personal Health Metrics Data*", IBM Journal of Research and Development*,

62(1), 1-1,2018.

[16] Ovunc Kocabas, Tolga Soyata, and Mehmet K. Aktas," Emerging Security Mecha- nisms for Medical Cyber Physical Systems*", IEEE/ACM Transaction on computa- tional Biology Bioinformatics,* Vol. 13, No. 3, June 2016

[17] Sergiu C,Gianpiero,"Practical Privacy Preserving Medical Diagnosis using Homomor- phic Encryption",*IEEE 9th International conference on cloud Computing*,2016

[18] Ovunc k,Tolga S,"Utilising Homomorphic Encryption to implement secure and private

medical cloud computing",*IEEE International conference on Cloud Computing* ,2015

Performance Evaluation of Adversarial Examples on Deep Neural Network

Architectures

Aruna A. Pavate[1], Rajesh Bansode[2]
Thakur College of Engineering, Mumbai, India
arunaapavate@gmail.com[1] , rajesh.bansode1977@gmail.com[2]

Abstract – **Tremendous advancement in machine learning over the last few years lead to use of deep neural networks in many applications from weather prediction to safety critical applications like disease diagnosis. Latest work observed that deep neural networks are vulnerable to the perturbed input samples as known as adversarial examples, very small changes considering single pixel can cause neural network models to make mistakes in predicting the output, this growth in safety concern of deployment of applications in safety critical environment. It is observed that the adversarial examples are transferred from one neural network model to another neural network model with considering adversary knowledge either black box which is similar in a real-life scenario with the adversary not aware of knowledge of the neural model to be attacked, white box or gray box. Adversarial attacks can be categorized by various characteristics. This paper provides a good inclination of adversarial examples in the context of computer vision and provides detailed discussion on different types of adversarial attacks on various types of neural network architectures and also elaborate the different metrics applied to validate the system performance. We conclude that adversarial learning is a factual threat to application of machine learning not only in the physical world but also during training the model and testing the model. There are some certain countermeasures exits but none of them can act as an individualistic solution for all the challenges. It remains an extensive challenge for the machine learning community to deal with robustness.**

Keywords:Deep neural network,adversarial attack, machine learning,computer vision

I. INTRODUCTION

Machine learning system gained remarkable success in almost all the applications. Machine learning used in many safety critical applications including self driving cars[1] [2], robots[3][4],Natural Language processing[5],speech recognition[6][7][8][9], malware classification[10][11],anomaly detection[12] and many more . Most machine learning algorithms are considered to map from an input domain to an output domain by detecting sample pairs of inputs and outputs from these domains. Lately machine learning algorithms are known to be vulnerable to adversarial attacks. In computer vision algorithms object detection, object segmentation, object classification is known to be vulnerable to adversarial attacks, by adding small perturbations to the input. In recent times the number of approaches have been implemented and examined to produce adversarial examples for different neural networks architectures and updating the parameters of the model to apply optimization techniques to find the optimal solution[13][14][15][16][17][18][19]. Szegedy et al. were first introduced the concept of adversarial samples , applied over image classification and misled the neural network with high confidence[13]. Deep neural network is noticed as blackbox as neural network works well with limited knowledge [20][21].

In this paper, we have explored the different attack approaches for creating adversarial examples. Most of the adversarial examples are created across the computer vision domain . we emphasis on one pixel attack as a one of the important adversarial attacks that demands very few adversarial information to deceive the various network architectures because of the immanent properties of the evolutionary algorithm and generates the low cost adversarial attacks to evaluate the robustness of the classifier based on Differential Evolution(DE). Using the available framework , implemented an experimental attack on image classification models using CIFAR-10 , CIFAR-100 data sets.

1.1 Discussion About Basic Terminologies

In this section, we highlight some major terminologies used during the work that are related to the generation of adversarial examples on Deep neural networks in the Computer Vision field.

Informal framework

General Framework for adding perturbation to the input sample that helps to generate the adversarial examples

Consider a Classifier K(X)= $\hat{Y} \in \{1, \ldots, N\}$ Let L(x) be the labels that belong to the class and M(X) be the logits.

$$arg\ max\ Li(X)_{i\epsilon\{1,\dots,N\}} = \hat{Y} \quad (1)$$

$$\sum_{i=1}^{N} L_i(X) = 1 \quad (2)$$

$$L(X) = softmax(M(X)) \quad (3)$$

Adversarial Example

Definition: An adversarial example is a sample input to the model such that perturbed input sample x' addressed at L(X') ≠ L(X) and the distance between || X'- X || being narrow i.e. sometimes unnoticeable to human eyes. Figure 1 represents the generation of adversarial sample.

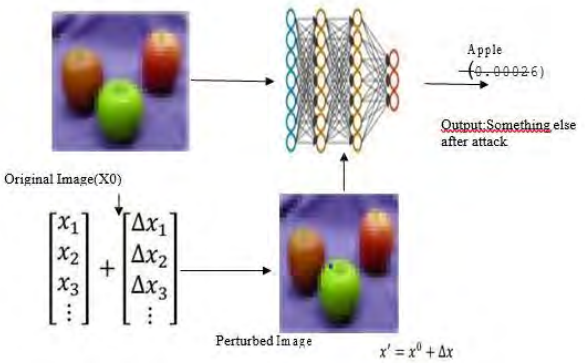


Fig. 1 An illustration of adversarial examples developed on Lenet. The predicted label belongs to something else than the actual class after the attack. The prior confidence was 0.00026. The actual class is an Apple. Here perturbed image x' is the smallest modification of an image by changing single pixel. X0 is the original image.

Threat Models:

Practically, the main objective of an adversarial attack is to alter the input sample as small as possible of a classifier in such a way that is not correctly classified. The attacker's knowledge can be considered as 1. Black- box 2. White box 3. Grey box

Black - box: The adversary has access to the input data and observers the output. A typical strategy of the attack is to use a proxy model for crafting an adversarial attack and adversarial samples tested on different architectures[22]. Another method is zero order optimization to generate adversarial samples without accessing model gradients[23]

White box : Here , the adversary is aware about the neural network model , network architecture type , number of hidden layers, activation functions , hyperparameters [15].

Grey box: Any variation between the white box and black-box attack is considered as grey - box attack where adversary has access to the architecture, training [24].

Most of the adversarial attacks are white box attacks where advisory knows all the details, but these attacks are can be converted into black- box models due to the transferability property of the adversarial examples[22]

Attacks Specificity:

Attacks can be either targeted or untargeted

Targeted attacks manipulate deep neural networks to misclassify the input to a specific class.Targeted attacks mostly applied on multiclass classification problem. E.g. Here Model misclassified an 'Apple' sample to ' bowl' class.

Non-targeted Attack: The output of the model can be represented arbitrary i.e. output does not assigned to specific class. These types of attacks are easy to craft and apply. E.g. Model misclassified an 'Apple' to as any other class

Attack Frequency : Attack can be one time or iterative . One time attack is generated only single time whereas iterative attack can take many attempts to generate adversarial examples. For some rigorous algorithms (reinforcement learning) , one time attacks are more suitable. Iterative attacks need many queries to be posted to success the attack so the computational cost is more to generate them as compared to single time attack.

Attack Measures : There are many measures useful for evaluating the performance in terms of efficiency and effectiveness of the proposed system. The following measures utilized here to evaluate system performance.

1. *Confidence:* This measure specify the intermediate probability of the labels of the expected class output to the successfully altered class label of an input sample from expected to target.
2. *Success Rate:* This measure specify that the observed probability of the original image that can be successfully changed to another targeted attack and non-targeted attack by modifying the original image
3. *Average Distortion:* Average distortion for a single pixel is computed by considering average modification of the channels (R,G,B) to evaluate the cost of the attack. If the value of the average distortion is

eminent such that it can be easily visible to the human eyes then the cost is high.

II. RELATED WORK

This section highlights different approaches for generating adversarial attacks. In the recent studies there are many approaches in the below described defeated by countermeasures. The presence of these approaches needs to be studied which improves the transferability and thus robustness of the neural network models.Szegedy et al. proposed box-constrained L-BFGS adversarial examples. Author discovered that mapping of input to output in deep neural networks interrupted to some extent. It is possible to cause the network to present an image belonging to another class by making certain perceptible perturbation which can be possible by exaggerating the network's prediction error. Also it is possible that the same perturbation can cause other different networks , trained using different dataset to analyze the same image belonging to a different class. The proposed concept applied on different networks like AlexNet,QuocNet,FC [15].

The Fast Gradient Sign Method(FGSM)[15] is an efficient solution tried for both targeted and nontargeted label attack. Fast gradient sign method computes the gradient cost function corresponding to the input of the deep neural networks.

S. Moosavi et al. proposed an accurate and simple method which computes the distance by adding small perturbation to the input sample to provide robust classifier. Deepfool computes the nearest distance from the input sample to the decision boundary of the perturbed input. The attack is applied iteratively to avoid the problem of nonlinearity in high dimension space. Deepfool proved on both binary and multiclass classifiers[17].

Papernot et al. proposed Jacobian-based Saliency Map Attack (JSMA) which is estimated as an expensive method as it runs much slower than the fast gradient sign method where Jacobian matrix computed for the input sample. The adversarial success rate 97 % achieved by applying small perturbation of 4.02% input features per sample[18].

Tom B. Brown et al. introduced a targeted universal attack using image patch and tried over the real world problem without knowing the knowledge of the type of classifier, scene- independent, angle of the camera etc.Trained patch retrieved using Expectation over Transformation(EOT) framework[25] where the expectation is applied over locations, random images and transformations. The attack evaluated across different Imagenet models like resnet50,inceptionv3, xception, VGG16, and VGG19. Here attack is examined and shown that small perturbations are not sufficient, large , local can also be possible to break classifiers. Confidencance 99%[26].

Universal perturbations were developed in [27] for targeted attacks by changing each pixel in the input image with high probability but the attacks were not tried in the physical world. This attack is an iterative attack that presents observations on the geometry of the decision boundaries of neural network.

Pin-Yu Chen et al. [28] proposed an elastic Net attack by introducing set of adversarial examples with small L1 bias and proved the similar performance like the other advanced methods in vaiours attack frameworks like FGM,C&W attack, IFGM. The developed attack enhanced transferability of attack and the capacity of adversarial training for the deep neural network.

Wieland Brendel et al.[29] proposed real world stable attack by considering adversary knowledge as blackbox. The implementation is based on boundary between adversarial and non-adversarial samples,and needs final class prediction. This attack is simple, does not require a substitute model, flexible in terms of adversarial criteria, and requires limited hyperparameter tuning .Author claimed that the decision boundary attack is more robust than gradient based attacks,score based attacks and transfer based attack.

Nicholas Carlini et al. proved that,providing defensive distillation doesnot improve the robustness of the classifier by developing new attacks for L0,L2 and L∞ which are much more powerful than the previous one.These attacks are successful with 100% transferability success rate between the secured model and unsecured model. These attacks were evaluated on three different datasets:MNIST,digit-recognition task, and CIFAR-10[30].

III. METHODOLOGY

In computer vision research, adversarial examples are original images that are input to the deep neural network that have been specially crafted to miss classify the input by the model. It is proved that deep neural networks can be fooled by an attacker by simply modifying the color of a single pixel [31] this approach is referred to as " one pixel attack". In many cases attackers can even cause the network model to generate the output of what they want i.e. targeted attack. We demonstrate this work into three different directions 1) Appling one pixel attack on CIFAR 100 dataset 2) Generation of the low cost adversarial attacks to evaluate the robustness of the classifier based on Differential Evolution(DE) 3) Analysis of one pixel attack on different network models. Our implementation is based on [31] uses a basic Convolutional Neural Network as shown in Figure 2.

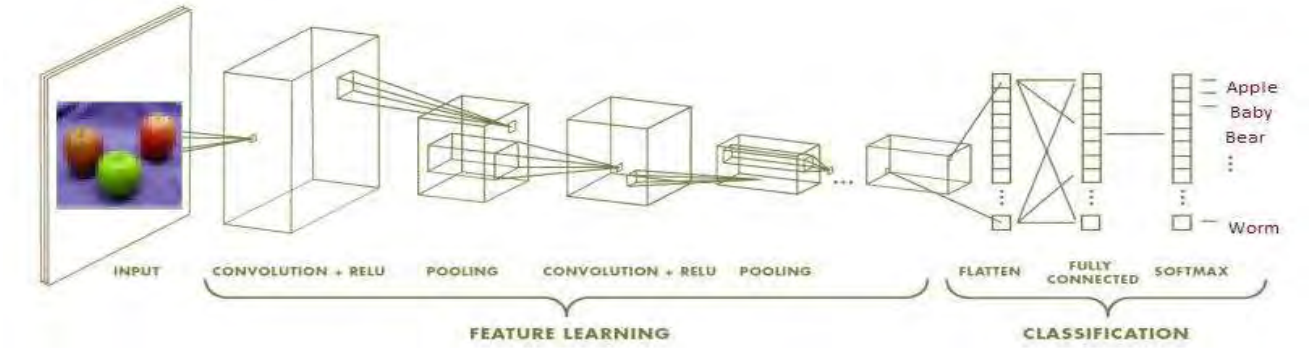


Fig. 2. Basic Architecture of Convolutional Neural Network applied on [CIFAR 100 dataset [32]]. The output of the convolutional Neural Network belongs to one of the class label among 100 labels.

One pixel Attack:

J. Su, et al.[31] proposed the approach that causes neural networks to misclassify an input sample by just crafting single pixel in the input sample. The attack is developed using genetic algorithms without crossover operation and is referred to as Differential Evolution (DE). The attack is generated for both targeted, untargeted attacks,Image specific which perturb single pixel to cause a misclassification . Our implementation is based on convolutional neural networks with different architectures.

Algorithm 1: One_Pixel in Pseudocode

```
1.  procedure perturb_image( (x,y,R,G,B), img):
2.    for x,img in zip((x,y,R,G,B), imgs): do
3.  perturbation pixels: pixels = np.split(x, len(x) // 5)
4.      for pixel in pixels:do
5.      At each pixel's x,y position, assign its rgb value:
                x_pos,  y_pos, *rgb = pixel
6.       img[x_pos, y_pos] = rgb
7.       return imgs
```

Input samples are encoded into vectors of five elements(x,y,R,G,B). Here x,y indicates the position of an image and the next three values represent the channel for the image. Here in this case R,G,B i.e. color value. Attacks are developed using differential evolution.**Differential Evolution:**

To find out the pixel which helps in providing successful attack, the problem is considered as optimization problem. Here the problem is either minimization of confidence in case of non targeted attack and maximization problem in case of targeted attack Differential evolution is a type of stochastic algorithm useful for global optimization problems.

Algorithm 1: Differential Evolution

1. Initialize population with random positions in search space :popsize=400
2. Until termination : adequate fitness reached do
3. For each solution in the population do:
4. Xs= x0+0.5(x1-x2)
5. Calculate fitness of each child if fitness[Parent]<fitness[child]:do
6. Replace: fitness[Parents]=fitness[child]
7. Repeat 100x , reating the updated solution as the parent solution
8. best= trial_denorm
9. Yield best, fitness[best_idx] : Best fitness is the "winner"

Targeted Attack: The main objective of the targeted attack is to minimize the confidence of the correct class and maximize the average of the probabilities of the other classes. This type of attack is very hard to implement as constrain the classifier to predict the wrong class than that of the actual one. There is no guarantee of success of attack , after several iterations the confidence of the classifier increases. Figure 3:

shows that letnet model to classify an image of an apple to an image of aquarium_fish

Fig. 3 Targeted Attack :One pixel attack implemented on trained on CIFAR-100 different types of DNNs :The right side image represents targeted attack for single sample with confidence 0.4634. The predicted output is aquarium_fish whereas the image belongs to Apple class

Untargeted Attack:

The objective of the Non Targeted attack is to classify an input sample to as given target class.To design a successful non targeted attack , the input sample is perturbed as to maximize the probability of the selected class. For CIFAR 10 and CIFAR100 we evaluated randomly drawn 100 samples from the validation set.

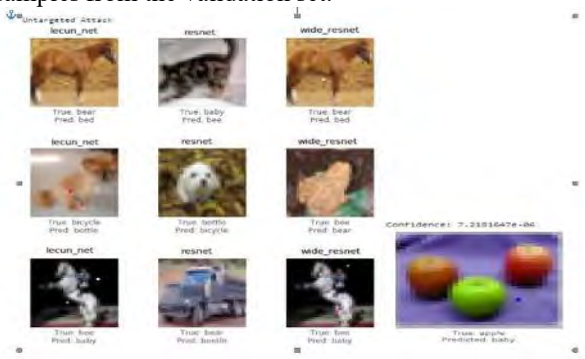


Fig. 4. Untargeted Attack :One pixel attack implemented on trained on CIFAR-100 different types of models of DNNs : The right image indicates Non-Targeted attack for single sample with confidence 7.21. True class is an apple whereas the predicted class belongs to baby

Fig. 5. Shows confidence of the model changes after adding small perturbation into the image . The middle image shows the success of attack on a single image for untargeted attack prior confidence 9.6221505e-05 & the attack successful

IV. RESULT AND ANALYSIS

We demonstrate the evaluation of one pixel attack with an original CIFAR-10 test dataset [31]. We have used 100 random images for non-targeted attack and 100 random images for targeted attack. It has been observed that the target DNNs have higher accuracy and confidence. All settings are kept the same as in the implementation of attack on CIFAR-10 dataset. The results of the attack shown in Table 1. The limitation of this attack is it works finest on low- resolution images. After implementing one pixel attack on various models on 100 different randomly selected samples it successfully generates100 adversarial images and calculated the success ratio. Multiple

experiments run by varying the number of pixels(here 1,3, 5) and the results were analyzed. The attack showed on CIFAR 10 works with an average accuracy of 88.08% and on CIFAR 100 with an average accuracy of 91% .

Table 1. Attacks on different CNN Models with 1,3,5 pixels perturbation

			CIFAR [0][3][]			CIFAR 100		
Model	parameters	Pixels	test accuracy	attack success (untargeted)	attack success (targeted)	test accuracy	attack success (untargeted)	attack success (targeted)
LeNet	62K	1	74.9%	63.0	34.4	96%	65.0	40.4
		3		92.0	64.4		92.0	64.4
		5		93.0	64.4		95.0	70.4
Pure CNN	1.4M	1	88.8%	13.0	6.67	81%	15.0	7.67
		3		58.0	13.3		50.0	18.3
		5		63.0	18.9		68.0	18.9
Network in Network	970 K	1	90.8%	34.0	10.0	95%	38.0	19.0
		3		73.0	24.4		77.0	27.4
		5		73.0	31.1		77.0	34.1
ResNet	470 K	1	92.3%	34.0	14.4	100%	39.0	34.4
		3		79.0	21.1		87.0	25.1

		5		79.0	22.2		78.0	25.2
Dense Net	**850 K**	1	94.7%	31.0	4.44	95.6%	33.0	4.44
		3		71.0	23.3		75..0	34.3
		5		69.0	28.9		76.0	34.9
Wide ResNet	**11M**	1	95.3%	19.0	1.11	96.3%	20.0	21.11
		3		58.0	18.9		45.0	28.9
		5		65	22.2		55.3	23.2
CapsNet	**12M**	1	79.8%	19.0	0.00	78.3%	22.0	3.00
		3		39.0	4.44		19.0	4.44
		5		36.0	4.44		35.0	5.44

V. CONCLUSION

The final aim of this work is to propose an attack over the Convolutional neural network. The applied attack shows that CNN is sensitive to adversarial attacks generated using stochastic differential evolution algorithm of optimization problem. It has been observed that the Convolutional neural network is susceptible to low cost and low dimension attack. The characteristics of neural networks have a major impact on the attack both in terms of success rate and the average distortion. The accuracy of the CNN models are not robustly related with the attempt of implementing a successful attack on an input sample. Perhaps it is most robust to the type of attacks. It has been observed that the CapsNet has the minimum attack success rate as compared to all other models both in case of CIFAR10 and CIFAR 100 dataset however the model is still vulnerable to attacks.

References:

1. Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., Et Al. End To End Learning For Self-driving Cars. Arxiv Preprint Arxiv:1604.07316 (2016).
2. Bourzac, K. Bringing Big Neural Networks To Self-driving Cars, Smartphones, And Drones.Http://Spectrum.Ieee.Org/Computing/Embedded-systems/Bringing-big-neural-networks-to-selfdriving-cars- smartphones-and-drones, 2016.
3. Janglova´ , D. Neural Networks In Mobile Robot Motion. Cutting Edge Robotics 1, 1 (2005), 243
4. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Et Al. Human-level Control Through Deep Reinforcement Learning. Nature 518, 7540 (2015), 529–533
5. Andor, D., Alberti, C., Weiss, D., Severyn, A., Presta, A., Ganchev, K., Petrov, S., And Collins, M. Globally Normalized Transition-based Neural Networks. Arxiv Preprint Arxiv:1603.06042 (2016)
6. Graves, A., Mohamed, A.-r., And Hinton, G. Speech Recognition With Deep Recurrent Neural Networks. In 2013 Ieee International Conference On Acoustics, Speech And Signal Processing (2013), Ieee, Pp. 6645–6649 7. Hinton, G., Deng, L., Yu, D., Dahl, G., Rahman Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., And Kingsbury, B. Deep Neural Networks For Acoustic Modeling In Speech Recognition. Signal Processing Magazine (2012).
8. N. Carlini, P. Mishra, T. Vaidya, Y. Zhang, M. Sherr, C. Shields, D. Wagner, And W. Zhou, "Hidden Voice Commands." In Usenix Security Symposium, 2016, Pp. 513–530.
9. G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, And W. Xu, "Dolphinatack: Inaudible Voice Commands," Arxiv Preprint Arxiv:1708.09537, 2017
10. Dahl, G. E., Stokes, J. W., Deng, L., And Yu, D. Large-scale Malware Classification Using Random Projections And Neural Networks. In 2013 Ieee International Conference On Acoustics, Speech And Signal Processing (2013), Ieee, Pp. 3422–3426
11. Yuan, Z., Lu, Y., Wang, Z., And Xue, Y. Droid-sec: Deep Learning In Android Malware Detection. In Acm Sigcomm Computer Communication Review (2014), Vol. 44, Acm, Pp. 371–372.
12. Chandola, V., Banerjee, A., And Kumar, V. Anomaly Detection: A Survey. Acm Computing Surveys (Csur) 41, 3 (2009), 15
13. C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow And R. Fergus, Intriguing Properties Of Neural Networks, Arxiv Preprint, Arxiv:1312.6199, 2013
14. Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, And Rob Fergus. Intriguing Properties Of Neural Networks. Arxiv Preprint Arxiv:1312.6199, 2013
15. Ian J Goodfellow, Jonathon Shlens, And Christian Szegedy. Explaining And Harnessing Adversarial Examples. Arxiv Preprint Arxiv:1412.6572, 2014.
16. Alexey Kurakin, Ian Goodfellow, And Samy Bengio. Adversarial Examples In The Physical World. In Iclr'2017

Workshop, 2016.
17. S. Moosavi-dezfooli, A. Fawzi And P. Frossard, Deepfool: A Simple And Accurate Method To Fool Deep Neural
Networks, Proc. Of The Ieee Conference On Computer Vision And Pattern Recognition, Pp.2574-2582, 2016
18. N. Papernot, P. Mcdaniel, S. Jha, M. Fredrikson, Z. B. Celik And A. Swami, The Limitations Of Deep Learning In Adversarial Settings, The 1st Ieee European Symposium On Security And Privacy (Euros&P), Saarbrucken, Germany, Pp.372-387, 2016
19. Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, And Adrian Vladu. Towards Deep Learning
Models Resistant To Adversarial Attacks. Arxiv Preprint Arxiv:1706.06083, 2017.
20. W. Knight, "The Dark Secret At The Heart Of Ai," Mit Technology Review, 2017.
21. D. Castelvecchi, "Can We Open The Black Box Of Ai?" Nature News, Vol. 538, No. 7623, P. 20, 2016.
22. Z. Zhao, D. Dua, And S. Singh, "Generating Natural Adversarial Examples," Arxiv Preprint Arxiv:1710.11342, 2017.
23. P.-y. Chen, H. Zhang, Y. Sharma, J. Yi, And C.-j. Hsieh, "Zoo: Zeroth Order Optimization Based Black-box Attacks To
Deep Neural Networks Without Training Substitute Models," Arxiv Preprint Arxiv:1708.03999,2017
24. Vivek B.S.,Konda Reddy Mopuri, R. Venkatesh Babu,"Gray-box Adversarial Training" ,Arxiv:1808.01753v1 [Cs.Cv] 6
Aug 2018
25. A. Athalye, L. Engstrom, A. Ilyas, And K. Kwok. Synthesizing Robust Adversarial Examples. Arxiv Preprint
Arxiv:1707.07397, 2017
26. Tom B. Brown, Dandelion Mané, Aurko Roy, Martín Abadi, Justin Gilmer,Adversarial Patch,Arxiv:1712.09665v2 [Cs.Cv]
17 May 2018
27. S.-m. Moosavi-dezfooli, A. Fawzi, O. Fawzi, And P. Frossard. Universal Adversarial Perturbations. Arxiv Preprint
Arxiv:1610.08401, 2016
28. Pin-yu Chen1∗ , Yash Sharma2∗ †, Huan Zhang3† , Jinfeng Yi4‡ , Cho-jui Hsieh3,Ead: Elastic-net ,Attacks To Deep
Neural Networks Via Adversarial Examples, Arxiv:1709.04114v3 [Stat.Ml] 10 Feb 2018
29. Wieland Brendel∗ , Jonas Rauber∗ & Matthias Bethge,Decision-based Adversarial Attacks: Reliable Attacks Against
Black-box Machine Learning Models,Arxiv:1712.04248v2 [Stat.Ml] 16 Feb 2018
30. Nicholas Carlini David Wagner University Of California, Berkeley,Towards Evaluating The Robustness Of Neural
Networks ,Arxiv:1608.04644v2 [Cs.Cr] 22 Mar 2017
31. J. Su, D. V. Vargas, And S. Kouichi, "One Pixel Attack For Fooling Deep Neural Networks," Arxiv Preprint
Arxiv:1710.08864, 2017
32. https://www.cs.toronto.edu/~kriz/cifar-100-python.tar.gz

Iot based application device for mentally challenged patients

Namdeo Badhe , Khushali Tank, Riya Rana, Saloni Verma' Sheetal Yadav
Information Technology, Thakur College of Engineering and Technology, India

Abstract: **Census 2001 has revealed that over 21 million people in India as suffering from one or the other kind of disability. This is equivalent to 2.1% of the population. Among the total disabled in the country, 12.6 million are males and 9.3 million are females. Although the number of disabled is more in rural and urban areas. Such proportion of the disabled by sex in rural and urban areas. Such proportion has been reported between 57-58 percent for males and 42-43 percent females. The disability rate (number of disabled per 100,000 populations) for the country as whole works out to 2130. This is 2,369in the case of males and 1,874 in the case of females. The treatment for mentally challenged patients can be very troublesome and difficult for the doctors and the patient's family as they have to be attended to constantly which is quite tiresome for the patient's family and doctors. Also, the patients may be unable to express any type of discomfort that they may be feeling, therefore, we create a device and supporting application to help locate the patient and monitor the patient's heart rate. The problem that the patient's family and doctors face is that they have to constantly attend the patient which is not always possible, even when they are not attending the patients they are always worried about the patient. They can never let go off the patient. This process is quite tiring and can be harmful for both patient and others if not done properly. Hence, there is a need for a device that can constantly track the patient and his condition. The device will monitor the patient's heart rate constantly so if there is any drastic variation in the patient's heart rate a list of emergency contacts can be notified. The device is connected directly to Web site so that the doctor and patients can easily track the patients' data.**

Keywords: IOT, RFID, IBAMP

I. INTRODUCTION

The Internet of Things (IOT) is the network of physical devices, vehicles and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data. Each thing is uniquely identifiable through its embedded computing system, but is able to interpolate within the existing internet infrastructure. Sensors should be installed to gather the data from surrounding of the patient and GPS tracking module must be implemented to trace the patient's location. Tracking system for mentally challenged people is an application which is primarily aimed to help mentally challenged people by continuously monitoring their location and notifying the emergency contacts in case some unusual behavior is seen. Addressing the people with issue with mental illness through single aiding system. In this project we are creating an environment for mental patients so that they can stimulate it. All these activities would be co-ordinated by using Raspberry Pi . The mental patients, who would use this system, would be able to recognize the environment and would get aware of any danger around.

A radio frequency identification reader (RFID reader) is a device used to gather information from an RFID tag, which is used to track individual objects. Radio waves are used to transfer data from the tag to a reader.

II.. LITERATURE SURVEY

In the Paper [1] it refers to motion detection, raspberry pi video recorder. Here accuracy can be increased up to 99% and system sized is decreased. In Paper [2] The Angel Sensor then transmits this data via Bluetooth low energy (BLE) to the user's mobile device and an associated mobile application. but the problem is Bluetooth is short range. Refers to raspberry pi temperature, humidity measurement, camera WiFi.. In paper [3], Tracking is focused either entirely on indoor or entirely on outdoor by using different devices and techniques .By using multiple wireless devices we have increased the efficiency of fetching the location over a long distance.

III. PROPOSED WORK

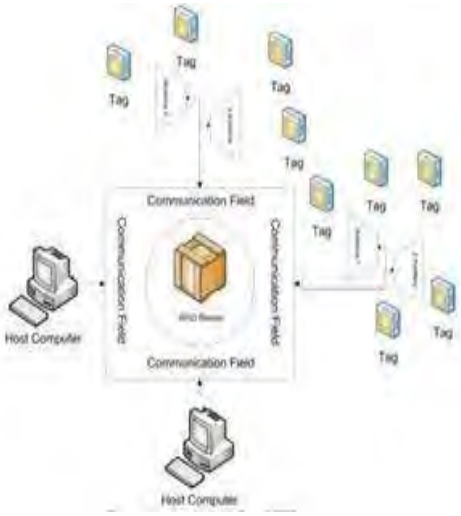
The proposed system will able to design a device which is a all-rounder i.e. it helps not only mentally challenged patients but also to the physically handicapped patients , the doctors, the guardians of the patient and other staff of the hospital to keep a track of the authentication, authorization, proper health and location of the patient. The device could be used in asylums, hospitals and also house. The RFID tags, RFID readers, temperature and humidity sensors, LED light, e-tracker, UNO R3

Development for Arduino and GPS Tracker which would help us in developing an environment where the patient, guardians and doctors would be able to keep track of health, location and authenticate and authorize themselves to use the facilities provided.

Besides we are also using Touch screen LED display which would act as a server for controlling all the operations mentioned above.

IV. METHODOLOGY

The idea is to use RFID chips embedded in, say the wrist of the patient, and as the patient passes through any dangerous kind of situation such as fire accident or facing any kind of obstacle in the environment, the chip is scanned by RFID readers mounted on either end. The tag ID is then sent to the server via an on board Wi-Fi module. The corresponding data is fetched from



the database, identification parameters are mapped, the status of the patient is detected and the situation of the patient is sent to their guardian or parents via text messages. All this should happen in "real-time", as the patient passes through the dangerous situation. The aim of our proposed innovation is to provide a Implantation of Access Control System, and take care of the mentally challenged person.

In our project, two levels of users are maintained
Administrator Level- owner/Guardian/Parent User Level- mentally challenged patients
In public places, we need to get the information about mentally challenged people in particular time and collect it by various means, like RFID (as part of Smart card based ticket), location aware mobile system, Wi-Fi membership) etc.
Multidimensional input data from any RFID system or other means as discussed, about its user, location,
current timestamp and RFID tag/card
After collecting and storing the current status of the patient immediately the status is sent to their guardian/parent so that they come to their ward and save them before any catstrophy takes place At the granular level, with the support of high computational power we even can able to track individual person movements.Fig.2 RFID Reader and tags Communication

V. RESULTS AND DISCUSSION

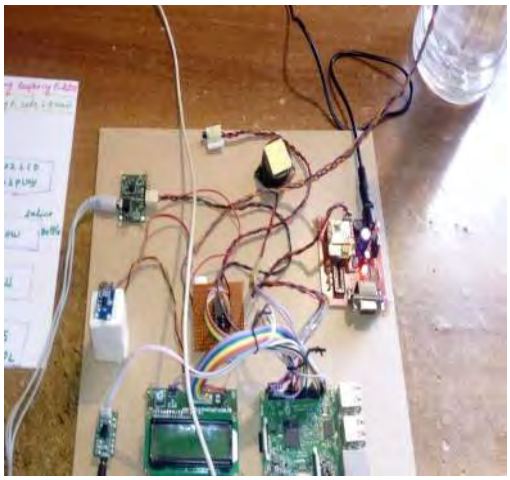


Fig 3 Working Model of the proposed system

VI. CONCLUSION AND FUTURE SCOPE

Thus by this, we conclude that the proposed device [Iot Based Application for Mental Patient (IBAMP)] and boost the mental patient's confidence by e- monitoring. It will also reduce the burden of guardian of patient.
In future, this device can be further used in multi-specialty hospitals.

VII.RESEARCH GRANT

Received Minor Research Grant by University of Mumbai of rs.30000/-

References

[1] H. Lamine and H. Abid, "Remote control of a domestic equipment from an Android application based on Raspberry pi card, " in Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2014 15th International Conference on, 2014, pp. 903-908.
[2] X. Wenchao, C. Ruizhi, C. Tianxing, K. Lei, Y. Yanqin, L. Xiao, et al., "A context detection approach using GPS module and emerging sensors in smartphone platform, " in Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2014, 2014, pp. 156-163.
[3] J. Chacon, H. Vargas, G. Farias, J. Sanchez, and S. Dormido, "EJS, JIL Server, and LabVIEW: An Architecture for Rapid Development of Remote Labs, " Learning Technologies, IEEE Transactions on, vol. 8, pp. 393-401, 2015.
[4] L. De La Torre, M. Guinaldo, R. Heradio, and S. Dormido, "The Ball and Beam System: A Case Study of Virtual and Remote Lab Enhancement With Moodle, " Industrial Informatics, IEEE Transactions on, vol. 11, pp. 934-945, 2015.

Heart Beat Measurement Using IoT

Sukesham Bissa, Gangadgar Chauhan, Nitin Bangar
Information Technology, Thakur College of Engineering and Technology, Kandivali Mumbai, India

Abstract: **There are many heartbeat sensors available in the market in watches and mobiles. We aim to present to you an affordable heartbeat monitor, which will help you keep your heart rates in check on a regular basis. Sometimes there is possibility of the monitor not giving accurate results due to some errors. We from our research found that the "Instant Heart Rate Monitor" received the highest accuracy and "Cardiograph" received the lowest. From this study we can easily conclude that the "Instant Heart Rate Monitor" is more accurate application when compared to the digital heart rate monitor and other such applications. Instant Heart Rate doesn't require heart rate straps and it monitors blood circulation with accurate heart health monitor. It detects the change in your finger nerves to provide accurate heart beat measurements. It measures your heartbeat in BPM accurately in less thna10 seconds. Heart rate can be affected by stress, workout, depression, emotions, training, fitness level, body composition, and medication use.**

I. INTRODUCTION

A heart rate monitor is a personalized monitoring device that allows individuals to measure their heart rate in real time and record the same for later study. It is mainly used by performers of various types of physical activities. Heart rate is the number of times the heart beats per minute and reflects different physiological conditions such as workload, stress and concentration on various tasks, drowsiness and the active state of the nervous system. The heart rate rises gradually during exercises and returns slowly to the normal value after exercise. Lower than the normal heart rates are usually a strong indication of a condition known as bradycardia, while the higher is known as tachycardia. Heart rate is simply measured by placing any finger over the person's arterial pulsation, and feeling, timing and counting the pulses normally in a 30 second period. Low-cost devices such as wrist watches are also available nowadays for the measurement of the heart rate. Such devices can give accurate measurements but their cost is usually as high as several hundred dollars, making them uneconomical. So this heart rate monitor with an optical sensor is a very useful instrument in knowing the pulse the patient.

II. MATERIAL AND METHODS

This project was carried out using patients body data.

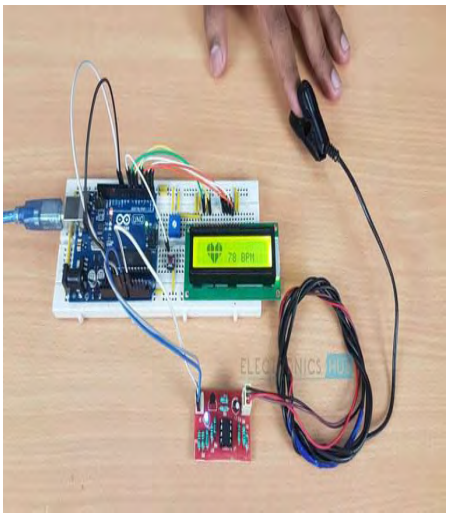
Study Duration: July 2018 to November 2018.

In this project are going to measure the heart beat using the pulse oximetry logic. The timer we will be setting for counting the heart beat is 30 seconds. There will be a set point we can decide, after 30s the heartbeat will be shown on the LCD. The code written in Arduino IDE takes the pulse of user and then simply divide it with the time constraint. After this calculation the pulse rate of user is shown on the LCD in beats per minute or simply BPM.

Resources Required: Programmers, Arduino. Software required: Thing Speak, Arduino IDE Hardware required: PC for development, hardware components for implementation.

III. RESULT

We were able to build a GUI for the web implementation using bootstrap. System is implemented using Django framework. The web has input interface where user will enter his parameters which we have mentioned. Our system will tell that whether user is diabetic or not.



The project intends to correctly analyze the input data on the basis of heart rate parameters to correctly judge the presence of any abnormality

IV. DISCUSSION

Thus, after the implementation of prototype we finally come to a conclusion and discussion that the cost of our project or system has to be reduced and efficiency has to

be improved as well. The main change to focus on is the size of the system. It has to be compact and easy to use in daily lives of users. So, in next phase our primary goal is to focus on size and efficiency of the proposed system.

V. CONCLUSION

- Currently the project focuses on gathering a structured authentic data.
- With reference to reviewed work we were able to find and define the problem definition we were able to find and define problem definition.
- We are clear with the idea of problem definition, algorithms, tools (front end, back end) required to implement the same.
- In the future, this project can be extended to give better health benefits and awareness to various classes and will be affordable for every middle-class person by using simple IoT and sensor technologies for sensing and sending data over the internet to a user friendly and device friendly websites.
- It can also be extended in health industry and hospitals as doctors will be able to monitor individual patients on regular basis and the risk of heart diseases can be easily reduced in such way.

References

[1]A PORTABLE WIRELESS SYSTEM FOR REAL-TIME ELECTROCARDIOGRAPHIC MONITORING
BY - You-De Liu, Yen-Chen Liu, Peng-Yu Chen, Sheng-Fu Liang, Kun-Chan Lan, Chung-Ping Young, Da-Wei Chang, Alvin W. Y. Su and Fu-Zen Shaw

[2] The Development of Microcontroller Based Low-Cost Heart Rate Counter for Health Care Systems
BY - Souvik Das

[3] HEART RATE MONITORING SYSTEM USING FINGER TIP THROUGH ARDUINO AND PROCESSING SOFTWARE
BY- Bandana Mallick, Ajit Kumar Patro

IOT based Smart Greenhouse

Amarjeet V. Singh, Rahul C. Yadav, Abhishek D. Singh, Neha Kapadia
Department Of Information Technology, Thakur College of and Technology
Mumbai, India
amarjeet952@gmail.com, rahulsince1997@gmail.com, abhishe216@gmail.com, neha.kapadia@thakureducation.org

Abstract – **In the recent era, with the rapid surge of growth in areas of automation and digitalization, our lives are improving as almost everything is undergoing mechanization which is replacing the primitive manual, intensive labour systems. Today, as Internet has become a vital integral part of our day-to-day lives, all the electronic devices also need to be brought on the network. This is the main aim of Internet of Things (IoT), which is one of the new upcoming technologies that helps in establishing connection, control, and management of standalone intelligent objects which are connected to Internet via different functions and protocols. In our project the main intention is to focus on incorporating IoT technology in the field of agriculture, specifically in the greenhouse system. Using several electronic components such as Intel Galileo, Arduino, and multiple sensors connected over the internet, the greenhouse parameters will be monitored and controlled with the help of different electronic devices like smart phone, computer etc.**

Keywords: Internet of things, sensor, automation, smart greenhouse.

I. INTRODUCTION

This proposed model of a Smart Greenhouse, helps the greenhouse workers to carry out the work in a greenhouse automatically without using much manpower. Greenhouse, a closed environment protects the plantations from harsh weather conditions and animal, pest attacks. The irrigatio n of smart greenhouse is done by the usage of mechanized drip irrigation, which operates on the parameter of soil moisture specified as optimal to the plants. Based on parameters from soil health card and data nitrate sensor proper amount of minerals can be applied by using drip irrigation techniques. Proper water management tanks are constructed and they are filled with water after measuring the current water level using an ultrasonic sensor. Plants are given the appropriate amount of light when it gets darker using adjustable led lights. The temperature and humidity sensor are used to monitor them**.**

The current architecture only supported the input data from sensors & display on web portal & only monitoring will not be efficient solution for maintaining Greenhouse, We propose monitoring and controlling the greenhouse components with the help of IOT. The current model water tank for irrigation, water tank is not a feasible solution our model uses drip irrigation. Most of the existing greenhouses have improper ventilation with lack of humidity control. We have proposed movable flaps which function for controlling the humidity & temperature of the greenhouse.

1.1 GREEN HOUSE

Even though India receives plenty amount of precipitation and has numerous large river systems, but one third of the total land for agriculture is attached via canal irrigation system. Rest majority of the portion relies on monsoon or tube-wells for the growth of plantation. There are places having excess water but face issues of land salinity, because of over irrigation and water-logging.

The Greenhouse, being a closed-frame structure safeguards
the plants from vigorous weather conditions such as hurricanes, Ultraviolet radiations, hailstorms, and insect & pest attacks. Here, the irrigation of agricultural field is carried out with the help of an automated drip irrigatio n process, which functions in accordance with the threshold soil moisture that is being set in such a way that an optimu m amount of water is given to the plants on a regular basis. Based on the data from soil health card, appropriate amount of nutrients, water and other requisites may be supplied by making use of drip irrigation techniques. Water management containers are also employed in the greenhouse, as they are filled with sufficient amount of water after determining the current level of water with the help of an ultrasonic sensor. The plants are provided with the requisite light wavelength in the form of glowing LED during the night time. Inside the greenhouse, air humidity and temperature are monitored by the humidity & temperature sensors respectively, and can be controlled with the help of movable roof flaps which function automatically as it reaches the threshold temperature and humidity.

1.2 BACKGROUND

In India, agriculture is still carried out in the traditional way and lags behind in utilizing the modern technologies available to mankind. This model aims to improve the preexisting agricultural practices by incorporating smart modern technologies for efficient and better production of greenhouse. The project provides a model of a smart greenhouse, which will aid the farmers in carrying out the farm work automatically, in an unmanned manner.

At present, there are numerous greenhouse models
which function in an automated manner, but most of them lack in minor parameters, which have an impact on the production of greenhouse. With the help of the proposed model, these shortcomings can be efficiently overcome, thereby improving the production of the greenhouse. Our project aims to bridge all these gaps which will help in efficient and less time of user being used in performin g redundant tasks. At present, with the advancement of technology there have been a lot of improvements in the greenhouse, but still there's ample room for further improvement to increase the efficiency and production rate of the greenhouse. There are greenhouses which have drip irrigation systeminstalled in them, sensors

that only collect data but are not programmed to perform any trigger- mechanism based actions, so we aim to overcome these shortcomings with the help of our proposed model. Production rate is fair in the existing greenhouse systems and can be boosted with much efficiency. Also, many of the greenhouses have been dependent on manpower or human assistance in some or the other way, which can be modified into a completely automated process by 24/7 Real-time monitoring of the green house.

1.3 IMPORTANCE

India being an agro-based country, agriculture is still being carried out in a traditional manner, and thus lags behind in the integration of modern techniques and technologies. About 57% population in India has been actively engaged in agricultural and associated activities, which make for about 15% of the country's GDP. Thus, it is very important in today's time for the involved stakeholders to leave behind the old conventional practices of agriculture and get ahead by modernizing the agricultural process with the help of better and efficient newer technologies.
The pecuniary contribution of agricultural activities to the
GDP of India is firmly decreasing with the country's wide- ranging growth in economy, although a large number of people are still persistently engaged in the agricultural sector.

1.4 OBJECTIVES & SCOPE

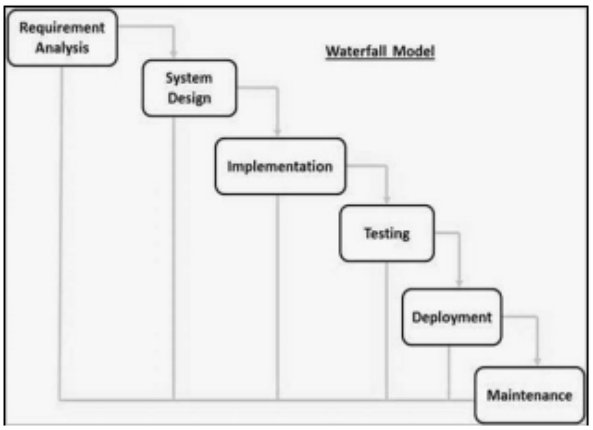
The greenhouse should be able to create multiple micro - climates with precision that favors custom condition for a variety of plants. The greenhouse will be fail-proof because all the components will be able to work independently so there is no single point of failure. The greenhouse should be able to conserve resources there by cutting costs of the plantation process and could lead to multiple growth cycles in a short span of time, increasing the productivity and revenue for the client.

The automated greenhouse in this project is meant for domestic use but can also be scaled to be implemented for bigger plantation. This project focuses on one plant only. However, this reference can be changed into any other plant.
The greenhouse has the ability to create a microclimate suitable for the growth of the plant,
which was achieved by controlling the factors such as humidity, temperature, moisture, light source.

To create a versatile greenhouse in which any conditions can be created just by altering soil properties, and using the IoT based monitoring and control to create other favourable conditions.
Wide area and safe monitoring available by authorised personnel.

II. METHODOLOGY USED

WATERFALL METHODOLOGY:



3.1 SUMMARY

Only monitoring will not be efficient solution for maintaining Greenhouse, We propose monitoring and controlling the greenhouse components with the help of IoT.A LAN based website would be area-limited. We can use a web portal & application which can be accessed from anywhere. Water tank is not a feasible solution; instead drip irrigation should be used. In earlier systems, there was improper ventilation with lack of humidity control. We propose movable flaps which function for controlling the Greenhouse humidity & temperature. Normal glass absorbs more heat. We can replace glass with acrylic material for efficient temperature management with the help of movable flaps using precise Servo Motors.

4.1 BLOCK DIAGRAM

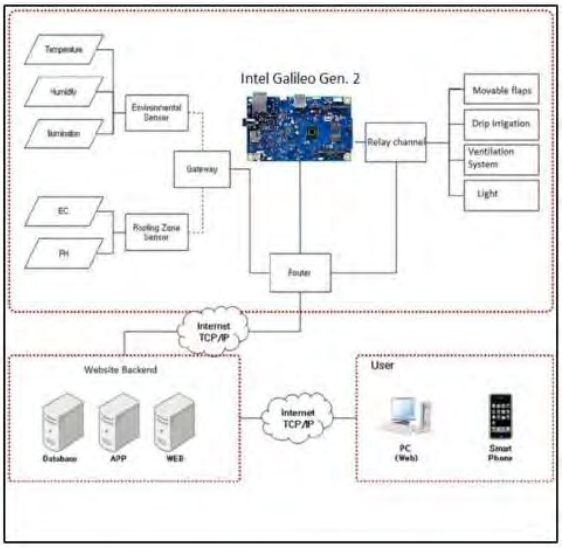


Fig. 4.1.1 Block Diagram

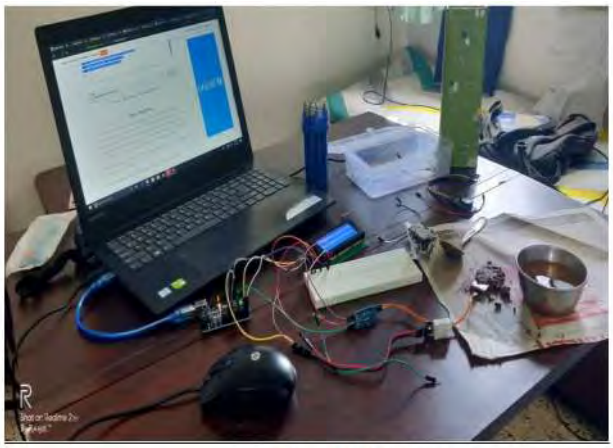


Fig. 4.1.2 Setup

cutting costs of the plantation process and could lead to multiple growth cycles in a short span of time, increasing the productivity and revenue for the client.

With the help of our proposed project, following features can be implemented and better results can thus be achieved:

Easy monitoring & control of the greenhouse through IoT.

Standalone operability of each & every sensor and modules.

Wide area accessibility with the help of web- based portal & applications.

Built-in movable flaps for proper ventilation of Greenhouse.

User-friendly interface of the web portal for monitoring the Greenhouse.

Advanced drip irrigation mechanism instead of conventional water tank method.

Adjustable light emitting diode (LED) lights for better growth of plant.

Fig. 5.1.1 Sensor Reading

The model not only overcomes the biggest challenge of deadlock situation but also increases the total efficiency of the whole object detection and obstacle avoiding system. Now assessing the Range of Application: The proposed model is capable of operating in military zones and areas where human presence is impractical, such as deep manholes for cleaning purposes, narrow places where movement is impossible for human. It can be an important asset to the government rescue programs and also oil industry.

III. CONCLUSION

Thus the results of an IoT based Smart Greenhouse allows precise position real time measurements and data transmission from the greenhouse to the interested farmer. Using the Intel Galileo and multiple sensors in an IoT based internet accessed agriculture application. All collected data can also be stored on a computer and day/time based reports and graphics are accessible for analysis at any time. The very specific data collection, easily adjustable sensor
The greenhouse should be able to create multiple micro - climates with precision that favors custom condition for a variety of plants. The greenhouse will be fail-proof because all the components will be able to work independently so there is no single point of failure.
The greenhouse should be able to conserve resources thereby positions and storage of data allows thorough analysis that can reveal even the most intricate trends in the farmin g process, thus directing the farmer how to improve the process for the next farming cycle. Furthermore, the long term collected data can be used by agriculture specialists to create more specific timetables and directions for growingspecific crops. Hence we use the actuators (motors/moving parts) as well, thus providing not only monitoring and data analysis but also precise control for the smart greenhouse.

References

[1] Shreyas Bhujbal, Yash Deshpande,Arpit Gupta, OjasBhelsekar "IOT Based Smart Greenhouse" IJIRSET, Vol.7, Issue 1, Jan 2018

[2] Prof. C.R. Dongarsane, Patil Pranav Balasaheb, Patil,Nilesh Rangrao, Patil Pranit Ramesh "Green House Automation Using IoT" IRJET, Vol. 4, Issue 1, Jan 2017.

[3] Veena Patil, Sonali B Patil, Swetha S Gogi, Satappa M Paramagond, Sidharta A Vardhamana "WSN BASED GREEN HOUSE OR IRRIGATION MONITORING AND CONTROL SYSTEM USING SMART PHONE" IRJET, Vol. 4, Issue 5, May 2017.

[4] Ravi Kishore Kodali, Vishal Jain, Sumit Karagwal"IoT based Smart Greenhouse" IEEE, 2016.

[5] Imran Bin Jafar, Kanij Raihana, SujanBhowmik, Shifur Rahman Shakil "Wireless monitoring system and controlling Software for Smart Greenhouse Management"

Internet of Things – Managing Projects

Vikrant Sankhe *Siemens Ltd.*
vikrant.sankhe@siemens.com

Abstract—**Excitement about invention and technology is omnipresent. Digitalization has made us believe in realities which were distant some years ago. Case in point, IOT. Imagine what life would be like when everything is connected. Where there is a surge in machine-to-machine communication making life simpler and to an extent more virtual, when everything from a streetlamp to a seaport becomes 'smart' in the true sense of the word, how every aspect of our lives and business will change? And who would make that possible? Internet of Things!**

The internet has made it possible for seamless collaborations across countries. IOT devices enhance the efficiency and effectiveness of communication and collaboration even more, as information and data is shared automatically between team members regardless of their location. In fact, it was one of the biggest predicted project management trends.

Keywords—IOT, OTA, Edge,

I. INTRODUCTION

Picture a client discussion scenario, "Just connect the device and get it on-line!" is often asked by the client manager while defining the use cases. During customer and user interactions, the primary IOT functions are defined: "We want to receive alarms, read-status, asset-control and verify the communication. All secure of course! Oh, by the way the project needs to be ready in 2 months and we need to have a demo next week for the investors." Does the situation sound all too familiar?

Now the value of IOT is not technology, but, lies in the creation of new propositions and potential revenue systems. The real value of IOT is at the intersection of gathering data and leveraging it. The information gathered isn't of much use unless there is an infrastructure in place, to analyze it in real time. Currently, the data used today is mostly for anomaly detection and control, not optimization and prediction, which provide the greatest value. This has resulted in a fundamental shift to approach and strategy. Intelligent and smart devices are ones that carry potential to source new products and services with high degree of efficiency.

The challenges associated with this digital transformation are significant. The vast new sums of data that these technologies generate will hold little value if suitable analytics techniques are not developed and implemented to inform intelligent decision-making across the enterprise. Moreover, to effect this transition companies will likely have to oversee and manage widespread changes relating to staffing, skills development, organizational structure and culture, and supplier relationships. This paper focuses on the challenges related to project implementation.

II. CHALLENGES

Some of the biggest challenges faced in the implementation of IOT are end-to-end services covering sensors, communication and application. Sensors measure, it evaluates; in short, it gathers data. To put things in perspective, the Internet of Things really comes together with the connection of sensors and machines. Coming back to the challenges faced, managing installations of sensors at field locations and managing these assets post implementation with respect to repairs, warranty, replacement etc. are crucial and should be always given serious thought and consideration.

III. THINGS TO REMEMBER

Often we ignore the power of the 'Platform' in a hurry to deploy 'Pilots'. The essence is missed out as the pilot calls for only few of the 'Things' to be connected. However, at the end state, there are millions of things that are connected along with associated multiple applications. Another big thing missed, is the firmware updates of these millions of devices in field. Hence, one needs to consider Firmware update OTA (Over The Air) functionality within integrated Platform.

As millions of devices get connected, the amount of data generated will also be BIG! Such massive amounts of unstructured data are of no use to anyone, unless it can be utilized well or has a context. To develop a model that forecasts behavior, data scientists require context and time-series data. Otherwise it becomes very difficult to consume this information and truly see what is happening now and what could happen in the future. People need real-time data to make the best possible decisions. With pervasive monitoring, this information is captured and delivered for business intelligence analysis. This BIG data also calls for the fundamental shift from on-premise to Cloud with Mobile integrated Applications.

One can make technology work efficiently in limited scaling in a 'point solution' with isolated business benefit. To scale up to an application having a positive effect on wider business performance requires a concrete plan backed with technological prowess. To be able to build on your initial success, consider the business strategy of your Internet of Things (IOT) project at the start. To effectively deploy IOT within this strategy, planning for robust network infrastructure is critical to be successful. Edge computing provides a means to assure reliability of the service to avoid unplanned downtime with high customer satisfaction. End-users at plants adopting IOT for near real-time predictive maintenance, process optimization and visibility also need a robust architecture with edge computing to assure high data integrity and avoid missed alerts. On the other

hand, the automatic transmission of a vast amount of data makes things automatically measurable and analyzable. This in turn, increases the efficiency of processes as decisions can be made quicker. The decisions will also be more accurate because they are based on empirical data. The biggest challenge of having so much data available is data security.

IV. SECURITY

IOT is all about physical "things", hackers that gain access cannot just perform the usual digital attacks like stealing data, moving money, or shutting down websites — they can cause havoc by tampering with infrastructure like electrical grids and traffic signals, or put lives at risk by meddling with healthcare devices, airplanes, or elevators.

In the recent past, Hackers used stolen credentials to gain access to the Ukrain Power Grid and cut power of 30 sub-stations and 225,000 customers. The attack included installation of custom firmware, deletion of files including master boot records, and shutting down of telephone communication.

In another instance, a team of researchers were able to take total control of a Jeep SUV using the vehicle's CAN bus. By exploiting a firmware update vulnerability, they hijacked the vehicle over the Sprint cellular network and discovered they could make it speed up, slow down and even veer off the road.

These examples are demonstration of emerging Internet of Things (or Threats) hacks: While companies often ignore the security of peripheral devices or networks, the consequences can be disastrous.

Security needs to be defined on 4 broad aspects:

1. Applications - Evaluate the specific threats w.r.t. applications. The risks associated with wearable fitness band are different to those facing a smart grid.
2. Integration - evaluate risk w.r.t. End-to-End integration. Security needs to be considered at various stages including, Platform, Communication, Sensor Security eg. Embedded malware and Third-Party Integration Security, especially connecting with third party solutions eg. Car data with Maps etc.
3. Availability - assess the impact due to non-availability of the security measures and
4. Confidentiality - the impact of losing out the confidential data.

We need to build IOT Solutions which are 'Secured by Design'. Plan for what to do if something goes wrong. Understand that compromise will happen, and plan a workflow for getting compromised credentials out of the system.

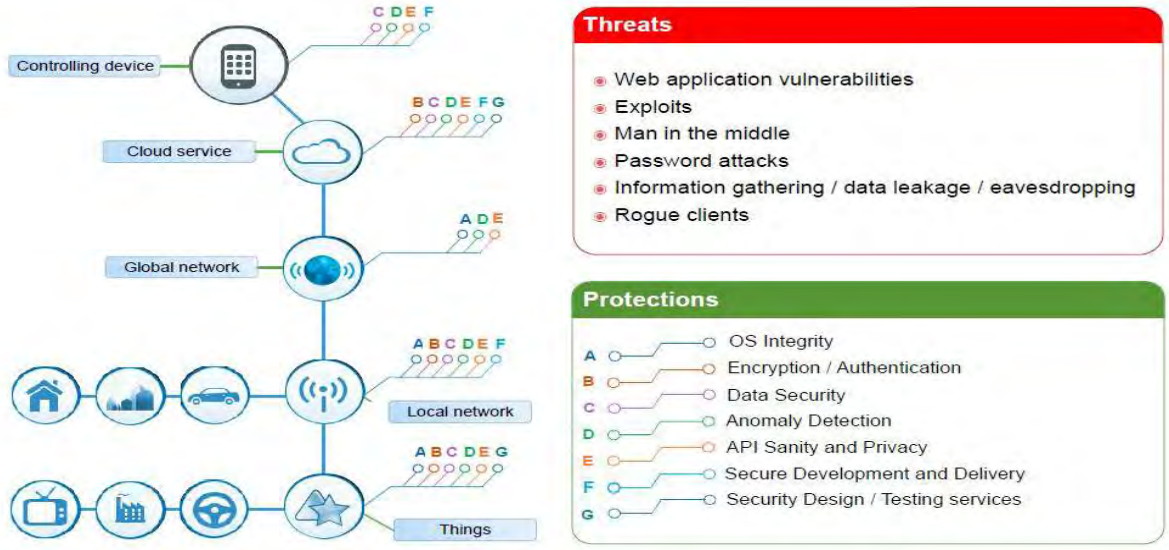


Fig. 1. Typical Security Architecture

V. ARCHITECTURE

We need a scalable Open Cloud Platform with data analytics and connectivity capabilities, tools for developers, applications and services covering end-to-end Security.

When we begin making things intelligent, it's going to be a major engine for creating new products and new services. We need to evaluate and utilize data and to gain breakthrough insights, drive the performance and optimization of assets for maximized uptime and also establish feedback for design improvisation as an ongoing process.

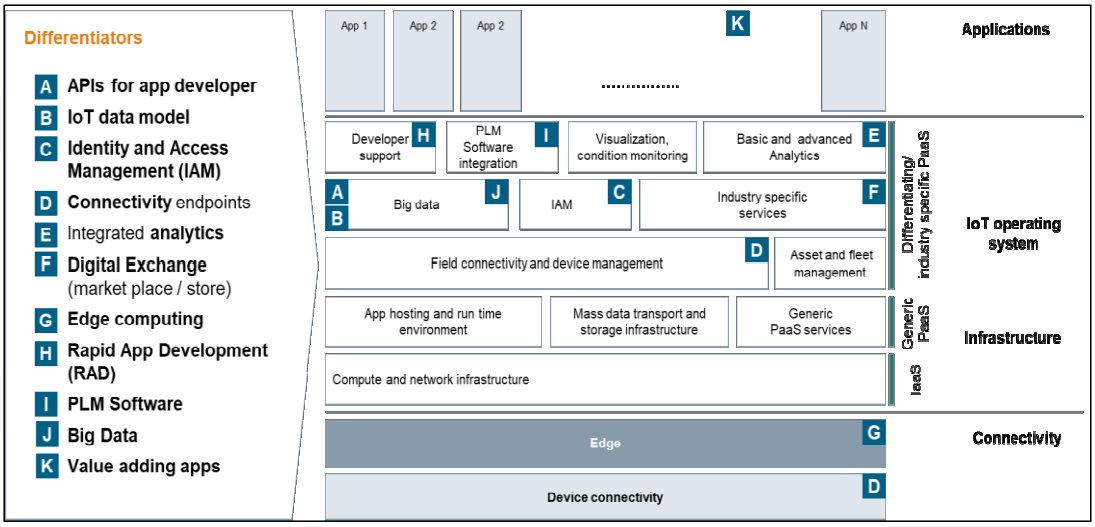


Fig. 2. IOT Architecture

VI. Conclusion

The Internet of Things is probably the biggest trend that's taking place in the world of technology as we speak. It is one that is going to give us the most disruption as well opportunity over the next few years. And hence managing the scale by design is utmost important.

U.S. economist Warren Bennis, who once quipped: "The factory of the future will have only two employees, a man and a dog. The man will be there to feed the dog. The dog will be there to keep the man from touching the equipment." Although intended as a joke, this statement certainly has a grain of truth. The fears and concerns that many people have in this regard are not unjustified.

References

[1] CIO Review India – Oct. 2017 Issue – 'Internet of Things – Managing Projects' by Vikrant Sankhe

[2] 'Internet of Threats' by Vikrant Sankhe (LinkedIn Blog) dtd. 7th January 2019

[3] Reference Architecture – Siemens MindSphere

Design of Automated Trash Collector Boat using CNN

Kiran Mungekar1, Bijith Marakarkandy2, Bhushan Nemade3
Thakur College of Engineering and Technology, Kandivali (E), Mumbai, Maharashtra 400101
mungekarkiran05@gmail.com, bijithm@hotmail.com, bnemade@gmail.com

Abstract – **Every day, a big amount of trash is released into the urban drainage which contributes to serious water pollution. Existing techniques are using a manual process for cleaning the trash which causes serious health issues and diseases. Existing trash collector boats are used only in big areas like river and sea for cleaning the surface water. The proposed system is sanitary and economical which automatically detects, collects and accumulates the floating trash using a conveying machine which can be controlled remotely from any location and this reduces human efforts. The proposed system is designed to collect trash from small and narrow space areas like drainage systems. The proposed system uses an augmentation technique for data preprocessing and Wireless sensor networks to control command remotely. It uses a convolution neural network for the detection of the floating trash. Once the floating trash is detected, the system calculates the position of the floating trash by the use of the high-resolution camera. The proposed system uses a CNN-based algorithm to extract image attributes and a multilayer perceptron's method to merge image attributes information to classify the floating trash. The proposed system is trained and validated against the physically labeled elements, expecting final classification accuracy more than 90% confidence under different testing approaches in real-time. Finally, a robotic conveying machine controlled by the microcontroller is used to pick up the floating trash and places it in the trash bin.**

Keywords: OpenCV, CNN, Deep Learning, IoT, Data Augmentation, ReLU Layer, Pooling Layer, Fully Connected Layer

I. INTRODUCTION

Water is the most precious gift to the humanity of nature and it is a very crucial element that supports life on earth. Today, the exponentially expanding number in the urban populace and boom of industries, water is experiencing the contamination issue. A big amount of floating trash is locating its manner into the urban drainage structures from where it's miles capable of travel through the city drain, streams, rivers, lakes, and estuaries till it subsequently reaches the open sea. Floating solids including refreshments/beverages can, plastic bottles, meals packaging, container, straws, and Styrofoam cups are a number of the most vital solid waste that may be found in the surface water, worsening pollution every day [1]. The undesirable presence of those wastes within the surface water will also contribute to environmental troubles, including drain logging and flash flood as well as potential diseases consisting of malaria and typhoid. On the other hand, for small streams and drainage, the manual cleaning method is frequently used. This scenario is undesirable since it can lead to health troubles and diseases to the operator due to the variety of impurities present in the water [2]. Hence, to overcome this situation a prototype robotic trash boat proposed system was expected to develop which can harvest floating trash from the urban drain and can be operated online/remotely.

Now a day, the trash collector boat has been widely used for cleaning of the surface water. Most of them are mostly targeted at massive place applications such as river and sea. The commercially accessible design of a trash collector boat calls for excessive manpower to function the system. In addition, the huge length of the trash collector boat makes it not possible for use in cleansing small regions consisting of small streams and drainage [3]. Therefore, a smaller length of trash collector boat has been designed to accumulate waste and floating solid from a slender and small place which includes small streams and drainage systems. The robotic trash boat proposed system was expected to design with automated detection, collection, and conveyor machine to accumulate the floating solids trash and can be remotely controlled from a distance. The proposed prototype is also portable, user-friendly, easy to understand and environmentally pleasant [4].

The objective of this proposed prototype is to design and build up a prototype of a transportable, robotic trash boat to collect trash from small and narrow space applications. The proposed system was conducted using open computer vision (OpenCV), convolutional neural network (CNN), deep learning and internet of things (IoT). In addition, it will be attracted and encourage the tourist to drive the boat across the surface water and make a contribution to keep it clean. The performance of the proposed system was examined and evaluated to determine the effectiveness of accumulating floating solids from surface water.

Fig. 1. Floating solids waste over the surface water.

II. RELATED THEORY

P S.H.Y.S. Abdullah et al. [1] the boat system works properly with the movements and speed corresponding to the collecting process of floating solids. But, the radio connection isn't strong that will bring about a connection loss sometimes at some point in the testing. It also wants to improve the quality of the image for better tracking of the trash collector boat. Mainly it offers the wireless remote control using radio waves, however, this results in connection loss in some scenarios, also needs the quality of the image for better monitoring and this problem can clear up by the usage of net connection, OpenCV, and convolution neural network and make it autonomous.

Sridhar Thiagarajan et al. [3] in this proposed system cups, trash was collected with the help of the classification algorithm. By using the fully connected neural network easy to identify the number of floating solids that provide more accuracy to the robot.

Abdoulaye Saadou-Yaye et al. [8] the compensation mechanism provided the ability to complete tasks in lower instances, but additionally gave the users enough control to perform finer movements. The compensation mechanism does have its boundaries, the system always continues to transport within the direction it was moving before the packet losses took place. To overcome this limitation, computer vision, object detection and object tracking used for determining the most likely direction to move in case of losses.

Shinji Fukui et al. [9] this paper proposed a new approach for object tracking. The tracker is primarily based on the particle filter. The object detector is used for monitoring the target object robustly. The proposed detector can come across similar objects with various sizes that exist near the target object. The positions of the similar objects are used in the procedure for calculating the likelihood and that for judging the situation under which the target object exists. Using statistical information makes the method robust tracking. The experimental results display that the proposed system can track stably the target objects. The problem is not about tracking stable objects, but the moving or floating solids over surface water live video streaming from the camera.

Prof. N.G.Jogi et al. [12] proposed a system targeted on modeling, design, and control of a pedal-operated boat, with emphasis on lightweight, portable appliances. Although this system able to acquire the rubbish from the lake with human intervention. To function the proposed system needs excessive manpower and this problem can be solved by using the solar panel the sufficient power can provide to the boat instead of pedal operating and Expending manpower. By making the prototype automatic using IoT and small enough to travel and collect trash from small and narrow space areas like drainage systems, also portable and lightweight.

III. PROPOSED METHODOLOGY

The proposed methodology is divided into two major systems I) Local System and II) Remote System as shown in Fig. 2. The identification of floating solids on the surface water could be accomplished from numerous points of view. This operation can be made seems easy by using image processing. The three-dimensional nature of the floating trash and the randomness in shape, size, and conditions of the floating trash and surrounding atmosphere situations like low light and their limitless possibilities, application of convolutional neural network with deep learning seems to be the only way out [1]-[3]. This would give the proposed system the ability to self-learn dependent on the experience they gain while they work.

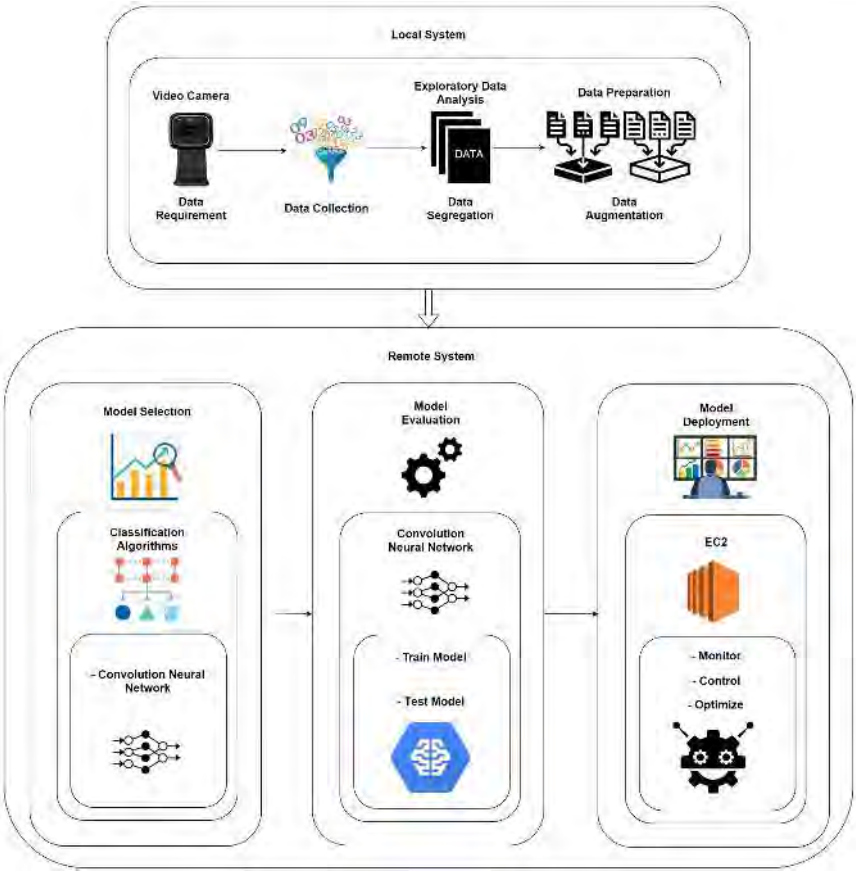


Fig. 2. Proposed system architecture.

1.1 Local System

There are billions of trash substances that are constantly being produced by the humans. The proposed system is meant for figuring out and clearing the seen floating solids. Examples of seeing floating solids include polystyrene blocks, plastic cups, bottles, and other plastic wastes, corn cobs, twigs, leaves, etc. are major concerns and for that, the "Data Collection" process plays an important role. OpenCV is a library of programming functions mainly used for image processing; the OpenCV read, write and display images functions/techniques are helps in data collection. This type of images was gathered at varying lighting conditions, some of them containing litter. It is further segregate into different labelled classes, this process is performed manually and called as "Data Segregation". Some of this labelled classes have many images while some classes have few images, which leads to imbalanced dataset. In the imbalanced dataset where some classes are contained more data as compared to other classes now at any given class the probability that this class is mathematically represented as follows using basic Bayes rule:

$$P(C0|x) = P(x|C0)*P(C0)/P(x) > P(x|C1)*P(C1)/P(x) = P(C1|x) \quad (1)$$

This is clearly show the effect of the priors and how it can lead to a situation where a class is always more likely than the other, so, that the classifier can't do a good prediction on the minority class. Now, in this kind of situation "Data Augmentation" techniques are used to generate more data and equalize the dataset [3]. But before applying data augmentation techniques each image must be resized and convert in grayscale form. The resize function from the OpenCV library is used to resize an image to the desired shape. The parameter here is the shape of the recently resized image. The images with a fixed-sized to train the model, take less processing power. In order to make a powerful floating trash classifier, augment the images via a number of random transformations, so that our model would never see the duplicate images. This has to be feasible through the "Image Data Generator" technique. Data augmentation is one approach to overcome overfitting, but it isn't sufficient since the augmented samples are still highly correlated. The model needs to concentrate on the most important features in the data, which is more relevant and helps to generalize the dataset in a better way. Finally, the proposed system used data augmentation to balance the dataset to improve the accuracy of the proposed model.

1.2 Remote System

After data preprocessing on the local system, the proposed system performs a model selection, model evaluation and model deployment task on the remote system. The "Model Selection" becomes the main task to identify which "Classification Algorithm" is fitting in the proposed system. The model selection task is done by using a "Cross-Validation" technique. The cross-validation technique is a simple and useful method for model selection and it is necessary to obtain an estimate of the error of model selection.

$$J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_\theta(x_{cv}^{(i)}) - y_{cv}^{(i)})^2 \quad (2)$$

This process is willing to take large processing power and also it is very time-consuming. At the "Model Evaluation" the "training and testing" phase is performed with "Convolution Neural Network" and the data of the number of images is passed through the following process:

Convolutional Neural Network, In the case of CNN the neuron in a layer, will only be connected to a small region of the layer before it, instead of all of the neurons in a fully-connected manner. Convolutional Neural Network (CNN, or ConvNet) is a sort of feed-forward artificial neural network in which the connectivity pattern among its neurons is inspired by using the organization of the animal visual cortex. The visible cortex has small regions of cells that are sensitive to unique areas of the visual field. Some individual neuronal cells in the brain respond (or fires) only in the presence of edges of a certain orientation. For example, some neurons fire when exposed to vertical edges and some when shown horizontal or diagonal edges. A computer understands an image using numbers at each pixel. CNN compares the image piece by piece. The pieces that it looks for are known as features. By finding rough feature matches, in roughly the same position in two images, CNN gets a lot better at seeing a similarity than whole-image matching schemes. The feature moves to every other position of the image and will see how the feature matches that area of the image.

$$h_w(x) = g\left(\sum_k w_{2,k,1}\, g\left(\sum_j w_{2,j,k}\, g\left(\sum_i w_{1,i,j}\, x_i\right)\right)\right) \quad (3)$$

ReLU Layer, Rectified Linear Unit (ReLU) transform function only activates a node if the input is above a positive quantity, while the input is below zero, the output is zero, however, whilst the input rises above a sure threshold, it has a linear relationship with the based variable. In this layer, remove every negative value from the filtered images and replaces them with zeros. This is done to avoid the values from summing up to zero.

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

Pooling Layer, In this layer the image shrinks stack into a smaller size.

$$C = [c1, c2, c3, c4, \ldots .cn\ ]$$

$$C_1 = max\{C\}$$

Fully Connected Layer, This is the very last layer where the actual classification happens. Now, the filtered and shrank images are put into a single list.

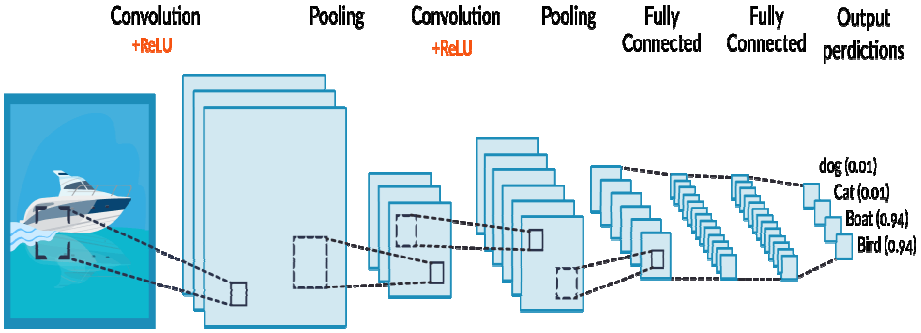


Fig. 3. Convolutional Neural Network.

Each prediction by way of the proposed system is compared with the manually classified label, and the "Confusion Matrix" helps to show the hits and misses of the classification of the proposed system. The system overall performance is evaluated with accuracy, precision, and recall. The accuracy of classification is defined as the percentage of images that are efficaciously classified:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} * 100\% \quad (6)$$

The precision represents the correctness of classification prediction of the proposed systems.

$$Precision = \frac{TP}{(TP+FP)} * 100\% \quad (7)$$

The recall represents the effectiveness of the classification prediction of the proposed systems.

$$Recall = \frac{TP}{(TP+FP)} * 100\%$$

After creating a trained model, it's time to load the model on the boat. The boat has a processor of its own that does all the detection and gives us information about the location and size of the object with the help of OpenCV and CNN. An image is just some number of pixels that get by using computer vision, it also gives the "x-axis, y-axis, height and width" of the identified object (floating solids in surface water). The identified object is marked in a square box and the value above the square box represents the confidence that the algorithm

determines the floating trash. With this information, it is easy to find the center point of the identified object.

$$Center(x,y) = \left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right) \quad (9)$$

This center point plays an important role to determine in which direction boat needs to move to come back to the vision center point and also set a threshold value that helps to control the speed and rotation of the motor.
The final fully tested proposed system is deployed on the cloud which helps to get the robotic vision with the visualization result of the analysis. The "Model Deployment" also helps to monitor, control and optimize the proposed system.

IV. PROPOSED ALGORITHM

Proposed System work in the following way:
Step 1: Get the data (the data in the form of images)
Step 2: Perform the segregation operation on the dataset
Step 3: Perform the aggregation operation on the segregated dataset
Step 4: Apply the classification algorithm for model selection
Step 5: Visualize the model accuracy
Step 6: Select the algorithm for model evaluation
Step 7: Split the dataset into the train and test set
Step 8: Train model on convolution neural network
Step 9: Test model on convolution neural network
Step 10: Predict the data from the boat
Step 11: Model deployment on cloud

V. RESULT AND DISCUSSION

The proposed system which is trained using an augmented image dataset is evaluated under the floating trash is placed with fixed and random orientations. The model performances are compared with a standard CNN model that takes the images as the input. The classification results are presented in Table 1.

Table 1. Expected result of confusion matrices for different classification models.

Evaluation Metrics	Standard CNN Model		Proposed CNN Model
	1st Test	2nd Test	Expected To Achieve
Accuracy (%)	87.7	80.0	83.85
Precision (%)	88.6	85.9	87.25
Recall (%)	96.8	89.2	93.00

The expected evaluation results show that the proposed system outperforms standard CNN model in terms of all three matrices (accuracy, precision, and recall), particularly for the floating trash.

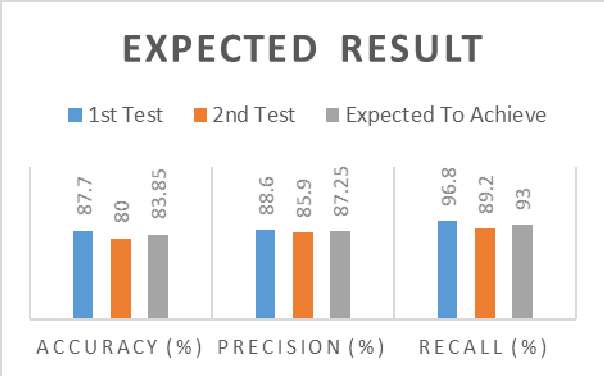


Fig. 4. Expected result of confusion matrices for different classification models.

VI. FUTURE WORK

To get, the more precious result, plan to include further classes of floating solid objects in our dataset and expect similar/better performance from the Deep Learning Algorithms. Also, try to segregate biodegradable and non-biodegradable floating waste using the Convolutional Neural Network.

VII. CONCLUSION

The computer vision combines with deep learning algorithms like the convolution neural network embedded and deployed on a robot, in order to successfully harvest floating trash from the urban drain.

References

[1]. S.H.Y.S. Abdullah, M.A.A Mohd Azizudin, A. Endut : Design and Prototype Development of Portable Trash Collector Boat for Small Stream Application, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-10, August 2019.

[2]. A. S. A. Kader,a,* M. K. M. Saleh,a M. R. Jalal,a O. O. Sulaiman,a W. N. W. Shamsuri : Design Of Rubbish Collecting System For Inland Waterways, Journal of Transport System Engineering 2:2 (2015) 1–13.

[3]. Sridhar Thiagarajan and G. Satheesh Kumar : Machine Learning for Beach Litter Detection, Springer Nature Singapore Pte Ltd. 2019, https://doi.org/10.1007/978-981-13-0923-6_22.

[4]. Jinqiang Bai, Shiguo Lian, Member, IEEE, Zhaoxiang Liu, Kai Wang, Dijun Liu : Deep Learning Based Robot for Automatically Picking up Garbage on the Grass, information: DOI 10.1109/TCE.2018.2859629, IEEE Transactions on Consumer Electronics.

[5]. Md. Nayamothullah, Mohammad Harun : Automatic control system by using the boat to remove garbage from the water, Department of Electrical & Electronics Engineering (EEE), December 15, 2018.

[6]. Jasper Princy Shinora W, Agilandeeswari L, Muralibabu K : An Adaptive Approach for Validation in Visual Object Tracking, Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15), doi: 10.1016/j.procs.2015.08.010.

[7]. Chakravartula Raghavachari, Aparna V, Chithira S, Vidhya Balasubramanian : A Comparative Study of Vision based Human Detection Techniques in People Counting Applications, Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15), doi: 10.1016/j.procs.2015.08.064.

[8]. Abdoulaye Saadou-Yaye, Julio Aráuz : Autonomy: Intelligent Command of Movable Objects, Peer-review under responsibility

of scientific committee of Missouri University of Science and Technology, doi: 10.1016/j.procs.2015.09.200.

[9]. Shinji Fukuia, Ryuji Nishiyamab, Yuji Iwahorib, M. K. Bhuyanc, Robert J. Woodhamd : Object Tracking with Improved Detector of Objects Similar to Target, Peer-review under responsibility of KES International, doi: 10.1016/j.procs.2015.08.227.

[10]. Sakuntala Mahapatra, Sumant Kumar Mohapatra, Biswa Ranjan Swain, Sushil Kumar Mahapatra : Simulation Based Algorithm for Tracking Multi-Moving Object Using Gaussian Filtering With Lucas-Kanade Approach, Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015), doi: 10.1016/j.procs.2015.04.216.

[11]. Resmi H.B, Deepambika V.A, M.Abdul Rahman : Symmetric Mask Wavelet Based Detection and Tracking of Moving Objects Using Variance Method, Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15), doi: 10.1016/j.procs.2015.08.012.

[12]. Prof. N.G.Jogi,Akash Dambhare, Kundan Golekar, Akshay Giri, Shubham Take : Efficient Lake Garbage Collector By Using Pedal Operated Boat, International Journal of Recent Trends in Engineering & Research (IJRTER), Volume 02, Issue 04; April - 2016 [ISSN: 2455-1457].

[13]. P.Hari Babu, Naveena, S.Sowmya, Ch.Rambabu, N.Prasanth : Implementing An Embedded Based Automatic Scarp Collector Bot In Waterbodies, International Journal of Advanced Research in Computer Science, ISSN No. 0976-5697, Volume 9, No. 3, May-June 2018, DOI: http://dx.doi.org/10.26483/ijarcs.v9i3.6022.

[14]. Sheikh Md Shahid Md Rafique, Dr. Akash Langde : Design And Fabrication Of River Cleaning Machine, IJSART - Volume 3 Issue 11 –NOVEMBER 2017.

Certificate Assurance through Block Chain

Sachin Sadare, Raja Choudhary
Founder & Director, Digital Dojo India, Mumbai, Manager - Digital, Digital Dojo, Mumbai India
sachin@digitaldojo.eu, raja@digitaldojo.eu

Abstract. **In the world today, there exists a need to mitigate fraudulent certificates and manipulation of records to ensure honest and seamless trade. Recently, modifying certificates and contracts for deceiving, misconstruing or fake anything is a commonplace happening. This is a clear breach of ethics. It takes our society to deeper darkness as a whole. Hence we need a system of issuing digital certificates on Blockchain technology that provides the following advantages - (1) leverage immutability of block chain technology to avoid certificate tampering, (2) public record mechanism to provide endorsement of the certification and (3) universal validation to prevent fraud and mutilation of records by involved entities. This paper presents the need, the solution, the high level architecture and the next research steps to make a digitally block chained certificate mechanism a universally accepted process.**

Keywords: digital assurance, digital certification, Certificate accreditation, Block chained certificates.

I. INTRODUCTION TO BLOCK CHAIN

To begin, we should discuss the historical backdrop of the Block chain. Before it was ever utilized in digital currency, it had humble beginnings as an idea in software engineering — especially, in the areas of cryptography and data structures. The specific primitive form of the block chain was the hash tree, otherwise called a Merkle tree. This data structure was protected by Ralph Merkle in 1979, and worked by checking and dealing with data between PC frameworks. In a shared system of PCs, approving data was essential to ensure nothing was adjusted or changed amid exchange. It additionally guaranteed that false data was not sent. Generally, it is utilized to keep up and demonstrate the respectability of information being shared.

In 1991, the Merkle tree was utilized to make a "secure chain of blocks" — a progression of data records, each associated with the one preceding it. The most current record in this chain would Contain the historical backdrop of the whole chain. And thus, the block chain was made. In 2008, Satoshi Nakamato conceptualized the conveyed block chain. It would contain a safe history of data exchange, use a shared system to time stamp and check each trade, and could be overseen self-governing without a focal specialist. This way, the block chain we know today was conceived, and the universe of cryptocurrencies. Block chain allows you to keep decentralized and distributed records of digital transactions. The first implementation took place in 2008 in the context of bitcoin as a digital currency, and although the Blockchain technology is no longer only in bitcoin, bitcoin is an example that will be used as paradigm of block chain. In bitcoin transaction occurs between anonymous users using public key cryptography. That means each user has its own private key that only he or she knows and one public key which will be shared with other users. All transactions are communicated to all nodes on the network. The nodes check transactions and group them in blocks. Each block is identified as a hash: a cryptographically unique value calculated on the content of the blocks and includes a reference to the hash from the previous block. So that blocks are linked. This chain of blocks is nothing but record of transactions or public accounting look shared by all the nodes in the network. In this way, all nodes can verify that the keys used are correct and the bitcoins transferred are from a previous transaction and they had not been spent already. However, transaction is only considered confirmed when it is part of a block added to chain. To add a block, it is important to undermine it, in other words, calculate its hash, which requires solving a unique mathematical problem of great difficulty that consumes some very considerable computing resources. As the power of the connected computer increases, the difficulty of the problem grows.

As a result, modifying its content is nothing but modifying its hash value, so link to the next block would fail and would break the chain, which, combined with the difficulty of repairing it, and with the fact that the rest of nodes have a copy of original string, makes the information contained in the block unalterable.

II. 2. Block Chain Decentralization

Blockchain is a public log or open registry that consequently records and confirms transactions. This open registry also known as the Distributed ledger technology (DLT) powers many virtual currency companies. Use cases for a transparent, verifiable register of transaction information is vast because of the fact that DLT works through a decentralized platform keeping it safe from fraud.

Fig. 1. Block chain Decentralization

The decentralization idea of block chain technology implies that it doesn't depend on central point for control. An absence of a single authority expert makes the framework more attractive and considerably more secure. The manner by which information is recorded onto a block chain encapsulates its most progressive quality: its value of decentralization. Rather than depending on a central authority to safely execute with different clients, block chain uses innovative consensus protocols over a network of nodes, to approve transactions and record information in a way that is incorruptible. As a block chain is a record of data it is critical that the data/information being stored is real and accurate. A consensus protocol is a set of rules that describe how the communication and transmission of data between electronic devices, such as nodes, works. Decentralization isn't to be mistaken for conventional distributed storage. Information which is stored on cloud is not directly stored on devices but kept on one central server somewhere else. Unlike block chain, it is very much centralized solutions. Information that is put away on the block chain isn't stored on a single server network, but rather distributed across wide range of devices over a P2P network.

III. Block Chain in Education

In all of the world's developed countries, the field of education is a central focus. This is due to the fact that the future of other important areas, including science, medicine, agriculture, industry and almost all others, is dependent on the country's education levels. Despite the remarkable developments in robotics and Internet and IT technologies, human resources are still the most precious resource to almost any business. Contrary to popular opinion, advanced technologies help to improve the sphere of personnel training and solve a lot of problems. Not only private but also in many state educational institutions are either getting ready to implement block chain-based tools, or they are conducting research to identify the areas where they need block chain implementations.

3.1 Problems in Education

Learning is no longer an activity that is performed during an initial period of formal mode, and which is enriched by experience in professional and vital practice. Lifelong learning has become a labor necessity; Human learning is a requirement of twenty- first century Citizenship which affects both the social framework in which we move and social learning. Formal education has soon transferred part of their programs to the temporary post-formal space; Post graduate studies and refresher courses became upgrade and specialization programs. The question that arises for years is how we can overcome the difference between formal education i.e. with accreditation and informal education i.e. without accreditation. However, accreditation of learning is not only complicated because of the diversification of the field from which training is offered, but it is increased by teaching methods and curricular elements which are materialized in an ideal pursuit personalized learning. The formal academic qualifications such as degrees and postgraduate, even with the correct documents it is not enough to major the capacity and knowledge of one person in particular subject. The kind of knowledge that we get in informal/non-formal education institute is appreciated and needed in industries or in everyday life. Due to this reason they are also counted in formal education when industries are recruiting. When industries are hiring they always asks for resume and if it is complemented with corresponding certificates then there are some challenges that industries will face to authenticate the documents and certificates. The traditional solution to for this problem is to contact the authority to cross validate the documents but this solution do not solve the problem of knowing the actual practical skills and knowledge acquired outside the formal education program in that particular subject.

3.2 Utilization of block Chain in Education Paper-based higher education diplomas, school certificates and extra training course certificates get lost and the original can be lost over time. Once put in block chain, the information about a student's grades and the courses that have been undertaken will not get lost, and it is impossible to change or falsify the information. At the same time, one can easily access this information with the owner's consent. Nowadays, the field of education is in need of central authorities to focus on checking certificates. Every day, educational institutions have to process thousands of requests from different educational organizations for sending copies of documents. They want to make sure the person has in fact graduated from a certain university. Putting information about education into block chain will remove doubt for employers about an employee's level of knowledge. Using block chain technology for storing information about one's level of education will make faking a diploma impossible. It is effective because it will solve the problem of low qualification levels of doctors, lawyers, engineers, economists and other popular occupations for which a formal education is required. A database can be established, which will include a person's skills, courses or training which have been completed, Assignments which are done and lectures which have been attended. Employers will be able to obtain and filter information about an employee's knowledge quickly and easily, with no doubt as to its relevance or reliability. Nowadays, recruitment for specific positions is a time-consuming and complex process. There is no single algorithm able to narrow the search down to only those applicants with the required knowledge. Rather than accumulating papers and certificates in education over their lifetime, people will be able to record information in block chain, which will make it available to employers anywhere in the world. In addition to the issuance of diplomas and certificates, educational institutions have to keep a paper record of students' progress for the entire duration of their study, thus creating a burden on the teaching and

administrative personnel as well as a lot of bureaucratic impediments. Non-official institute, Digital Dojo India provides Blockchain certificates for their courses which is a great technology to use for educational startups to gain trust from the people who are learning from them. This block chain certificates includes assignments, completion certificate and some other things about that particular person that will help the recruiter to know the knowledge of that person in the particular subjects. Block chain certificates or mostly known as Blockcerts consist of four basic components: Issuer or the institution that creates the certificates. Certificates adjusted with the requirements of the open badges initiative which contains a wide range of statements about skills, achievements or characteristics of the student, all recorded in a chain of blocks. Verifier is someone who will verify the certificates to see if it has been altered, and that it has been issued by a particular institution and belongs to particular individual, without having to depend on the distributor. Wallet is where each student can store their certificates and also share with others.

Fig.2. Block chain certificate structure.

IV. Critical View

The passion with which some sectors of education system have been pursuing block chain has raised alarms, both by the origin of the enthusiasm, as well as the consequences that an actual application of technology as we know it could have. As this time speaking of the consequences of an implementation is to move into the field of speculation as much as the literature that is promoting it. Despite this there are few challenges posed by the block chain in following four points:

(1) It is not about easy and immediate implementation for social, technological and economic reasons.
(2) It can lead to unacceptable consequences depending on the direction it is working and intentions with which it is made.
(3) It presents challenges in areas such as privacy, transparency, functionality and value of the certification, as well as official and public institutions.
(4) It concerns citizen who have seen in these years as some technological changes not always desirable. In coming few years, we will come to see if we are moving towards more economical and easier to use systems with this technology or not.

The three key elements of the imagined future in the discourse on block chain in education that deserves special attention:

(1) The anti-institutional disposition of Block chain. (2) Its dependence on decentralization that does not necessarily implicate democratizes, but sometimes means dismantling the public sector (3) The invocation of trust as a key social behavior mediated by technology.

V. Conclusion

Block chain is a disruptive technology that, after a few years of implementation as the basis of digital currency, is showing itself to be an open resource with possibilities in different field. The key to the interest in the technology lies in its ability to move from a system of centralized data logging to a distributed system that ensures no alteration of information and maintenance of privacy. However, concept of learning certification can also be applied in relation to training projects that strengthen the personalized itineraries of each student, and that at the same time, try to solve some basic problem such as fair and equitable grades or the accreditation of the progress made. It is likely that in coming year we see a growing interest in this technology.

References

1. William Mougayar, The Business Blockchain: Promise, Practice, and Application of the Next Internet Technology, Wiley (2016).
2. Melanie Swan, Blockchain: Blueprint for a New Economy, O'reilly 1st Edition, 2015.

Emerging Wireless Technologies: A Comparative Study

Sowmyashree, Chitvan Mehrotra
Thakur College of Engineering and Technology, Kandiviili (E), Mumbai, India

***Abstract*—With increasing demand for high speed wireless communications, many wireless technologies have emerged in the recent years. Wireless communications provide faster transmission of information, easy access without the need of cords, greater flexibility and mobility to the users. This paper contains comparative study of emerging wireless standards defined by Institute of Electronics and Electrical Engineers (IEEE) such as WLL, Wi-Fi, Wi-Max, Bluetooth, Li-Fi and Zigbee and GiFi technologies.**

Keywords: Emerging Wireless technologies, WLL, Wi-Fi, Bluetooth, Zigbee, Li-Fi, Wi-Max, Gi-Fi

I. INTRODUCTION

Wireless network is a network which uses wireless data connections between nodes within the network. In wireless networks devices can stay connected to the network but roam untethered to any wires. Whereas, the wired networks is used to establish a connection between two or more nodes or components at earlier time. It would work when an application needs short range communication containing smaller number of nodes. But when number of node increases or the range between them increases, the wired method becomes very untrustworthy, untidy and complex method. The main limitation of wired technology is mobility or portability, and the major reasons for this are need of power outlets to run them and the physical length of the wires which makes it a very bulky system and a really difficult manage it. Get damaged easily, so frequently there is a need for replacement. In addition, it is an expensive method and too much time consuming. To overcome limitations of the wired technology, the wireless communication technologies came into existence.

Wireless communication uses mediums like radio waves, microwaves, infrared waves, and satellite communications. Radio waves are the first wireless communication medium which permit a user to speak over short distances. Microwave wireless communication is an efficient sort of communication, mainly this transmission uses radio waves, and therefore the wavelengths of radio waves are measured in centimeters. Satellite communication is one sort of self-contained wireless communication technology, it's widely spread everywhere the planet to permit users to remain connected almost anywhere on the world. When the signal (a beam of modulated microwave) is sent near the satellite then, satellite amplifies the signal and sent it back to the antenna receiver which is located on the surface of the earth. Satellite communication contains two main components like the space segment and the ground segment. The ground segment consists of fixed or mobile transmission, reception and ancillary equipment and the space segment, which mainly is the satellite itself. Infrared wireless communication communicates information during a device or systems through IR radiation. IR is electromagnetic energy at a wavelength that's longer than that of red light. It is used for security control, TV remote and short range communications. In the spectrum, IR radiation lies between microwaves and visual light. So, they will be used as a source of communication

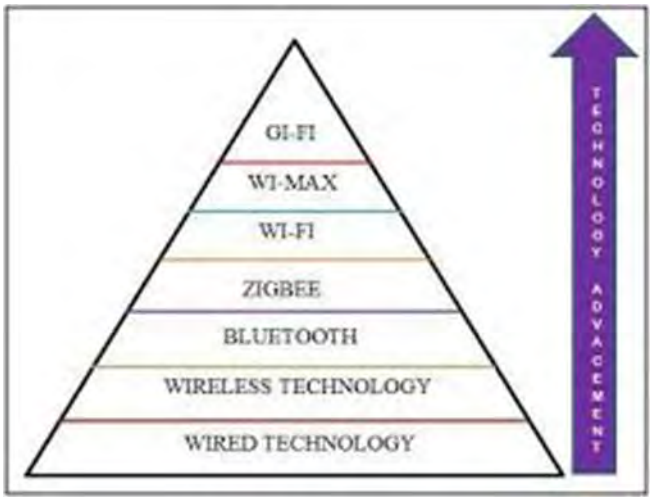


Figure 1: Network Evolution

Figure 1 shows technology advancement from wired network to Gi-Fi technology[9]. With the assistance of wireless technology we will send data over long distances with none quite trouble and which is more trustworthy and expedient compared to wired methodology. In industry, several sorts of devices, sensors, data collection devices and lots of more are connected in several ways via various protocols, if this is often done using wireless technology than it can replace all the physical limits of wires.

The advantages of wireless networks include increased efficiency, availability, flexibility and security. A variety of IEEE standards and protocols are used for wireless communication like Bluetooth, Zigbee, RF modules, infrared and lots of more. It is utilized in many areas like industrial, commercial, home use, health care, Smartphone, laptops and lots of other fields. The particular wireless standard is being chosen as per the need and specification of the merchandise or the appliance. For example, Bluetooth is employed in Smartphone, laptops, home automations, music players, etc. where several meters of range and data transfer in tens of MBs are required.

II. WIRELESS TECHNOLOGIES

A. WLL

Wireless local loop (WLL), is that the utilization of a wireless communications link because the "last mile / first mile" connection for delivering plain old telephone service (POTS) or Internet access (marketed under the term "broadband") to telecommunications customers. Wireless local loop (WLL) provides two-way calling services to the stationary or “fixed” users, which is meant to exchange its wireline counterpart [1]. WLL systems can be based on satellite based, cellular based, and microcellular based and fixed wireless access technologies.

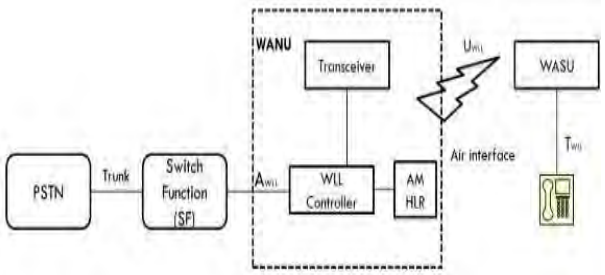


Figure 2: Typical architecture of WLL

The given architecture consists of three major components i.e WANU, WASU and SF. The WANU consists of various components which include several base stations transceivers or radio ports (RP), Radio port control unit, an Access manager (AM), an HLR. It provides various functionalities like authentication, Air interface privacy, over-the-air registration of subscriber units, Operations and Maintenance, Routing, Billing, Switching, functions, Transcoding of voice and data.

Wireless access subscriber unit (WASU) provides an air interface UWLL towards the network and a standard interface TWLL towards the subscriber. The power supply for it is provided locally. The interface includes protocol conversion and transcoding, authentication functions, signalling functions. The TWLL interface are often an RJ-11 or RJ-45 port. The UWLL interface are often AMPS, GSM, DECT then one.Switching Function (SF) is related to a switch which will be digital switch with or without Advanced Intelligent Network (AIN) capability, an ISDN switch or a Mobile Switching Centre (MSC). The AWLL interface between the WANU and therefore the SF are often ISDN-BRI or IS-634 or IS-653 or such variants.

Limitations of WLL involves, the technology is more costly due to the need for research and development. The technology has not been tested over an extended term of your time for reliability and repair costs. Certain technologies aren't available altogether areas, which leaves people with the unsupported technology disconnected. Wireless technology requires that data be sent over open space, which makes it vulnerable to interception and reduces the safety of the transmission.

B. Bluetooth

Bluetooth is a short range data communication platform. This helps in the connection of devices for different applications. It is not a perfect form among the latest, emerging wireless technologies while there's a requirement to transfer large amounts of knowledge. They usually carry the data at speeds of upto 721 Kbps. Bluetooth uses frequency hopping techniques with the carrier modulation using Gaussian Frequency Shift Keying (GFSK). [2][7].

Up to eight devices are often connected together using this technology. In this system one will act as a master and seven others as slaves. A network of devices connected together is named Piconet or PAN (Personal Area Network). Bluetooth uses unlicensed ISM band frequency of 2.45 gigahertz (GHz). Devices in Piconet are synchronized to a standard clock and frequency-hopping pattern, with the master device. In frequency hopping technique, the transmission frequencies are adjusted automatically by the paired device.

Spread spectrum frequency hopping uses frequency within the range of 2 .4 to 2.485 GHz. Signal is transmitted in 79 individual frequencies within the given range with a niche of 1MHz. Frequencies are randomly chosen and Bluetooth transmitters change frequencies 1600 times every second, such that, chances of two neighboring signals transmitting at same frequency is extremely less. Bluetooth 2.0 can manage a speed up to 3Mbps and is backward compatible with Bluetooth 1.0 devices, which features a transfer speed of 1 megabit per second (Mbps). Bluetooth connectivity is meant for devices within the range of about 10m radius. Bluetooth technology is employed in devices like headphones, speakers, wireless mouse, data exchange between mobile phones etc.

Figure 3 shows the architecture of bluetooth. The radio layer is responsible for Modulation/Demodulation of data for transmitting (OR) receiving over air. The baseband layer is liable for Controlling the physical links via radio assembling the packets and controlling frequency hopping. The link manager protocol controls and configures links to other devices. The host controller interface(HCI) handles communication between the host and therefore the module. For this purpose, it uses several HCI command packets like the even packets and data packets. The L2CAP layer converts the info obtained from higher layers into packets of various sizes. The RF COMM provides a serial interface with wireless application protocol (WAP) and object exchange(OBEX). WAP and OBEX provide interface to other communications protocols. The TCS(Telephone control protocol specification) provides telephony service. The SDP(Service discovery protocol) allows the devices to get the services available on another Bluetooth enabled device. The applications present within the application layer can extract the services of the lower layers by using one among the various profiles available.

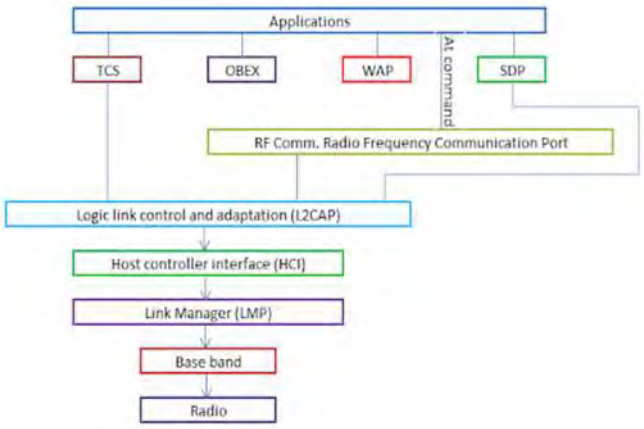


Fig 3: Architecture of Bluetooth

Advantages of bluetooth are, Bluetooth avoids interference from other wireless devices. It is easily upgradeable and consumes low power. Bluetooth provides better range than infrared. Bluetooth is employed for voice and data transfer. Bluetooth devices are available at very cheap cost. Free to use if the device is installed with Bluetooth. The technology is adopted in many products like head set, in car system, printer, web cam, GPS system, keyboard and mouse.

There are certain disadvantages like low bandwidth compared to Wi-Fii, losing connections in certain conditions and security issues.

C. ZigBee

ZigBee works in 2.4 GHz frequency. Basically it uses mesh, star and cluster topologies. They are inexpensive, general purpose, self-organizing and a robust wireless technology. Usually, ZigBee networks need a number and a coordinator for its operation. ZigBee is based on IEEE 802.15.4 standard. To achieve low power consumption, transmission distance is restricted to 10-100m distance. Also to transmit the data over long distances, we normally pass the data through a mesh network of intermediate devices. Commonly ZigBee is employed in applications that have long battery life and network security[7].

Some of the characteristics of ZigBee includes, Various transmission options, Security mechanisms, and operations in 2.4GHz frequency. ZigBee operates in the 915 MHz (Americas) and 868 MHz (Europe) and includes Power saving mechanisms. Z.igBee Operating over 16 channels. ZigBee works in 2.4 GHz frequency. Basically it uses mesh, star and cluster topologies. They are inexpensive, general purpose, self-organizing and a robust wireless technology.. ZigBee is based on IEEE 802.15.4 standard. To achieve low power consumption, transmission distance is restricted to 10-100m distance. Also to transmit the data over long distances, we normally pass the data through a mesh network of intermediate devices. Commonly ZigBee is used in applications that have long battery life and network security.

Working of ZigBee: The physical layer of ZigBee is supported by the 802.15.4 specification. The particular may be a bundle based radio convention planned for ease, battery-worked gadgets. The convention permits gadgets to impart during a mixture of system topologies and may have battery life enduring a couple of years. The Zigbee convention has been made and endorsed by part organizations of the Zigbee Alliance. In more than 300 heading semiconductor producers, engineering firms, OEMS and administration organizations contain the Zigbee Alliance enrollment. The Zigbee convention was intended to offer an easy to-utilize remote information result portrayed by secure, solid remote system architectures. The Zigbee convention is meant to impart information through unfriendly RF situations that are normal in business and modern applications.

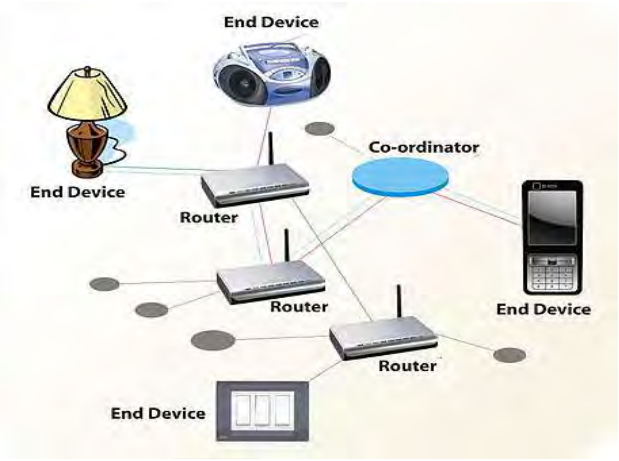


Fig 4: Operating modes of Zigbee

Figure 4 shows three operating modes of Zigbee: Zigbee router, Zigbee Coordinator(ZC) and Zigbee end device (ZED) Zigbee Router: this is often an intermediate format of application that acts sort of a data router, where the info are often transferred from one device to a different using the wireless networks.

Zigbee Coordinator (ZC): The ZC is recognized to the basis of this network tree because it acts a bit like a bridge for this suite with the opposite networks. Using ZC one can explore a plethora of data about network, which makes it the last word Trust Centre where users can get answers about their queries that are related to Zigbee.

Zigbee End Device (ZED): This is entitled with a set of characters that helps the user to connect to the parent node of the system (the parent node can be either the router or the coordinator). Only thing about ZED is that it doesn't have any functions for relying the info from other devices, but the inter-relationship of nodes allows you to get sleep mode where one can significantly increase the battery lifetime of the

device.

The Advantages of Zigbee is it is cheap and easy to utilize networks. Zigbee has a very long battery life. Zigbee is secured and used globally. Zigbee network supports >64000 nodes in network with low latency. The product, accessibility to broader markets, interoperability and vendor independence. Zigbee is employed in industrial quite environment and also for building automation. Zigbee has very long primary battery life, also dynamic and static mesh and star network.

Disadvantages of Zigbee is it cannot access in mobile phones because of its hardware size. Zigbee network only supports 20-250 bandwidth size. Zigbee only supports those application which has only short range area. Zigbee network supports large number of nodes. Zigbee stack size is very less [3] [4]

D. Wi-Fi

Wi-Fi (Wireless Fidelity) is a technology that allows many electronic devices to exchange data or connect to the internet wirelessly using radio waves. The Wi-Fi Alliance defines Wi-Fi devices as any "Wireless Local Area Network (WLAN) products that are supported the Institute of Electrical and Electronics Engineers' (IEEE) 802.11 standards. The key advantage of IEEE 802.11 devices is that they permit less costly deployment of Local Area Networks (LANs). For places where running cables to each device isn't practical, like outdoor areas and airports, they will host wireless LANs. Wi-Fi uses Wireless Adapter of a PC to translate data to radio signal and transmits it using antenna. Wireless router is used by Wii-Fi to receive and decode the signal. Also, it helps to send the information to Internet using physical and Ethernet connection. Figure 5 shows architecture of wiFi.

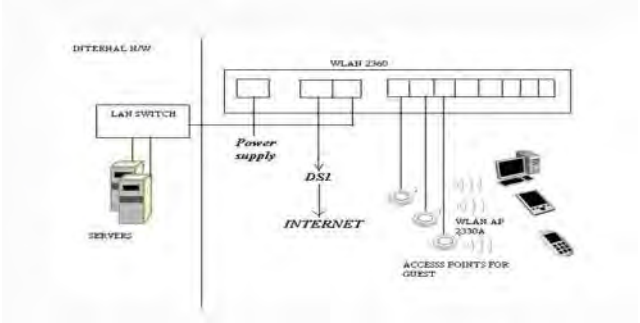


Figure 5: Architecture of WiFi

Wi-Fi uses very high frequency in comparison to other wireless communication methods, lies within the range of two .4 GHz or maybe higher frequency, such as, 5 GHz and thus the signal can carry more data. The 802.11 standard in which Wi-Fi is defined, has a few sub versions [7].

(a). 802.11a

Transmission frequency in 802.11a standard is 5 GHz. It improves the data speed up to 54 megabits per second. 802.11a standard uses orthogonal frequency-division multiplexing (OFDM). In this technology, radio wave is split into several sub-signals before they reach a receiver.

(b). 802.11b

This is the slowest and least expensive method. In 802.11b, the quality transmission frequency is 2.4 GHz and features a rate of 11 megabits per second.

(c). 802.11g

802.11g uses same transmission frequency as that of 802.11b, that is, 2.4 GHz. But its rate is extremely high since, it uses orthogonal frequency-division multiplexing (OFDM), that is, 54 megabits of knowledge per second.

(d). 802.11n

802.11n is the most widely used standard. It supports a data rate of 140 megabits per second. The technology is backward compatible with all the standards mentioned above.

Advantages of WiFi

• Convenience

The wireless nature of such networks allow users to access network resources from nearly any convenient location within their primary networking environment

(a home or office). With the increasing saturation of laptop-style computers, this is often particularly relevant.

• Mobility

With the emergence of public wireless networks, users can access the web even outside their normal work environment. Most chain coffee shops, for instance, offer their customers a wireless connection to the web at little or no cost.

• Productivity

Users connected to a wireless network can maintain an almost constant affiliation with their desired network as they move from place to put . For a business, this suggests that an employee can potentially be more productive as his or her work are often accomplished from any convenient location.

• Deployment

Initial setup of an infrastructure-based wireless network requires little quite one access point. Wired networks, on the opposite hand, have the extra cost and complexity of actual physical cables being run to numerous locations (which can even be impossible for hard-to-reach locations within a building).

• Expandability

Wireless networks can serve a suddenly-increased number of clients with the prevailing equipment. In a wired network, additional clients would require additional wiring.

• Cost

Wireless networking hardware is at the worst a modest increase from wired counterparts. This potentially increased cost is nearly always quite outweighed by the savings in cost and labor related to running physical cables.

Disadvantages of WiFi

• Security

To combat this consideration, wireless networks may prefer to utilize a number of the varied encryption technologies available. Some of the more commonly utilized encryption methods, however, are known to possess weaknesses that a fanatical adversary can compromise.

• Range
The typical range of a standard 802.11g network with standard equipment is on the order of tens of meters. While sufficient for a typical home, it'll be insufficient during a larger structure. To obtain additional range, repeaters or additional access points will need to be purchased. Costs for these items can add up quickly.
• Reliability
Like any frequency transmission, wireless networking signals are subject to a good sort of interference, also as complex propagation effects that are beyond the control of the network administrator.
• Speed
The speed on most wireless networks (typically 1-54 Mbps) is way slower than even the slowest common wired networks (100Mbps up to many Gbps). However, in specialized environments, the throughput of a wired network could be necessary.

D. WiMax

WiMAX (Worldwide Interoperability for Microwave Access) is gaining more popularity because it is faster and does not require any sort of wired connections[5]. WiMAX techonology provides 30 to 40 megabit-per-second data rates. It makes use of OFDM (Orthogonal Frequency Division Multiplexing) and MIMO (Multiple Input Multiple Output) technology for its function. WiMAX applications Provides mobile broadband connectivity, Telecommunications, Data, and ITPV services. WiMax can be used as a source of Internet connectivity and is better for Metering and Smart grids

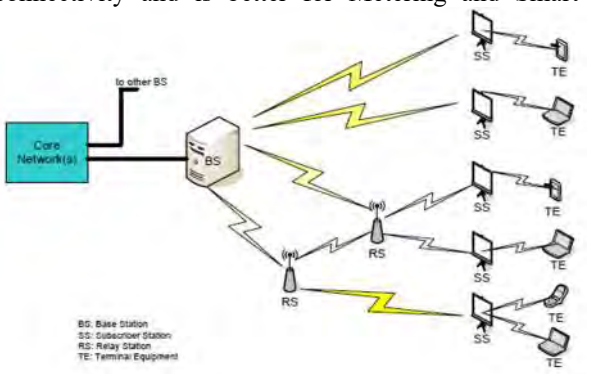


Figure 6: Wi-Max network architecture

WiMAX architecture [8] consists of two types of fixed (non mobile) stations i.e.Subscriber Stations (SS) and Base Stations(BS). Subscriber station serves as a building (business or residence) and Base station (BS) Provides connection to public network and provide Subscriber Station with first-mile access to public networks The communication path between SS and BS has two directions: Uplink direction (from SS to BS) and Downlink direction (from BS to SS)

Advantages of WiMax

A very high speed voice and data transferis supported over longer distances by WiMax. A Single WiMax BS serves hundreds of users. Using WiMax high speed can be achieved. Higher coverage range and cellular like performance are achieved with mobile wimax.

Disadvantages of WiMax

LOS (Line of Sight) connection is needed for subscribers located at the far distant from the WiMAX BS.Bad weather conditions such as rain will interrupt the wimax signal and often cause loss of connection.
Wimax is high power consuming technology and requires significant electrical support. High initial cost is required because it isn't backward compatible with any of the wireless cellular technologies..

E. Gi-Fi

GiFi (Gigabit wireless) is a wireless communication at data rate of about one million bits (gigabits) per second.GiFi is the first integrated transceiver integrated on one chip that permits transfer of videos and audios at a speed of 5 GB per second. GiFi is understood to possess ten times the present maximum wireless transfer rate during a range of ten meters [6]. GiFi as a wireless technology features a number of applications are often utilized in different places and devices that include wireless networks, smartphones, mm-waves and media access control. Due to high power consumption, low range frequency operation and slow rates of Wi-Fi and bluetooth, GiFi is used. GiFi as a wireless technology makes use of a really tiny antenna which will be employed by mobile devices and it's an excellent tool for wireless home equipment. GiFi aims at solving the low rate problems because it has an integrated receiver and transmitter on one chip. Gi-Fi provides better data transfer rate with minimum power consumption. GiFi offers the deployment of service within a few minutes, where other technologies may take hours.. GiFi wireless technology incorporates one subscriber station that's made available for various access points. It supports IEEE 802.15.3C standard millimeter wave wireless networks that are used majorly for communication between computer devices.

The subscriber station basically comprises of a little antenna that's mounted on the highest so as to support the sunshine of sight operations. In order to avoid any interference, it transmits multiples signals across the trail of transmission, at an equivalent time having different frequencies.

Working of GiFi.

Using time division duplex for receiving and transmission the data files are converted to RF 60GHz range by making use of two mixers from an IF range. The output is then stored during a power amplifier, that stores a millimeter wave antenna within. The RF incoming data is first converted to IF signal at 5GHz then converted to a traditional data range. A heterodyne construction is employed for this procedure to avoid any leakage due to direct conversion. With the supply

of a 7GHz spectrum, the info gets transferred during a matter of seconds. GiFi is small sized. The chip measures approximately 5mm on each side, incorporates a small antenna and makes use of a 60 GHz spectrum. Therefore, GiFi offers high mobility and portability. It also provides better coverage area. GiFi can be used in several applications like, household and office appliance, inter vehicle communications etc.

III. COMPARATIVE STUDY

Table 1 shows compassion of wireless technologies like, Buetooth, Zigbee, Wifi, WiMax anf Gi-fi with respect different parameters such as IEEE specification, frequency range, data transfer rate, coverage, cost, topology, type of modulation and application areas and amount of power consumption. We cannot draw a conclusion of which one is better, as superiority of these technologies are application dependent.

Table 1: Comparative study of emerging wireless standard

Features	**Bluetooth**	**Zigbee**	**WiFi**	**WiMax**	**GiFi**
IEEE specification	802.15.1	802.15.4	802.11a/b/g	802.16	802.15.3C
Frequency Range	2.4 GHz to 2.483 GHz	2.4 GHz	57 to 64 GHz	2.5 to 3.5Ghz/5.8Ghz	2.4 GHz and/or 5 GHz
Data transfer rate or speed	1Mbps	250kbps	11Mbps	1Ggbps	5 Ggbps
Operating range or coverage range	10 meters	10-100meters	100 meters	50Kmeters	10 meters
Cost	Low	Medium	Low	High	High
Operating frequency	2.4GHz	868/915MHZ ; 2.4 GHz	2.4GHz;5GHz	2.3-3.5GHz	57 to .64GHz
Network topology	Adhoc, P2P, star,Mesh	Mesh, star,cluster	Infrastructure, Ad Hoc,IP&P2P	Infrastructure IP network	Infrastructure; Ad Hoc,Mesh, star,cluster
Modulation type	FHSS/GFSK	BPSK; QPSK	DSSS	QAM	BPSK
Application	WPAN	WPAN	Wireless LAN, Internet	Wireless MAN, Internet	WPAN, inter vehicle communications
Power consumption	5mW	10.2mW	10mW	~5mW	<2mW

IV. CONCLUSION

In today's world, emerging wireless technologies are used in numerous applications with strong and faster information transmission. In this paper, overview of different wireless standards are discussed along with their advantages and limitations. Suitability of these standards depends on the applications. Comparative study of technologies is shown with respect to various parameters.

References

[1] Anthony R. Noerpel, Hughes Network Systems Yi-Bing Lin, Csie/Nctu, "Wireless Local Loop: Architecture, Echnologies and Serr/ices", IEEE Perqonal Communications 1070-9916/98 June 1998

[2] Kapil Bhoria, Harish Rohill, "A comparative study of emerging wireless standards: Bluetooth, WiFi, WiMax", International journal of applied engineering Research, ISSN 0973-4562,Vol 6, No.8, 2011

[3] Jobina Mary Varghese, Nibi K V,Vijo T Varghese And Sethuraman Rao, "A survey of the state of the art in Zigbee" International Journal On Cybernetic & Informatics(IJCI) Vol.4,No.2,April 2015.

[4] Abhishek Kumar, Sandeep Gupta ,"Study on ZIGBEE technology" International Journal Of Engineering Sciences& Research Technology (IJESRT),2013.

[5],Hadia Abd Elrahman Abdella Ali1, Dr.Mohamed Abaker Hussein, "Wireless Telecommunication Technologies: Li-Fi Vs. Wi-Max Vs. Wi-Fi Vs. ZigBee Vs. Bluetooth", International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 08; August - 2016 [ISSN: 2455-1457]

[6] Dr.Dinesh Vithalrao Rojatkar, Tejal Sanjay Shambharkar, " GIFI next generation wireless technology", January 2017, Volume 4, Issue 01 JETIR (ISSN-2349-5162) Journal of Emerging Technologies and Innovative Research (JETIR)

[7] https://en.wikipedia.org

[8] IEEE 802.16: WiMAX Overview, WiMAX Architecture Mojtaba Seyedzadegan and Mohamed OthmanInternational Journal of Computer Theory and Engineering, Vol. 5, No. 5, October 2013

[9]https://www.mepits.com/tutorial/253/wireless/latest-emerging-wireless-technologies

Internet of Things (IoT) and Radio Frequency Identification (RFID) Technology- A Survey

Mrs.Apeksha Waghmare
Thakur College of Engineering & Technology, TCET, Mumbai, India

Abstract— **The significance and futuristic deployment of the Internet of Things (IoT) are broadly reviewed and conveyed in this paper. The "Internet of things" (IoT) is fetching an increasingly growing topic of discussion everywhere around the world and has a huge impact on our lives. Radio Frequency Identification system (RFID) on the other hand is the usual technology that supports identifying objects through radio waves. This technology is called the Internet of Things (IoT). RFID is often seen as a prerequisite for the IoT. This paper introduces the technologies of RFID and IOT along with its applications and challenges.**

Keywords- Internet of Things (IoT), Radio Frequency Identification(RFID),Future Technologies

I. INTRODUCTION

In the IoT, environmental and existence things, additionally named "things", "objects", or "machines" are increased with computing and communication technology and be a part of the communication framework. In alternative words, the net of Things (IoT) is that the network of the physical objects or things embedded with natural philosophy, software, sensors, network property, that permits these objects to gather and exchange information. The interactions are tired varied ways in which i.e. machine to machine (M2M), person to person, person to machine, machine to person and then on. Connecting objects may be wireless, like the frequency identification (RFID), or sensing element radio technologies that provide, severally, identification of things and sensing of the setting. The affiliation could also be wired, like transmission line communication (PLC). PLC offers information transport over electrical media and has pioneered the in-home networking property of electronic client devices e.g. good fridges, smart TVs, good heaters, etc. In the IoT, distinguishing, sensing and mechanically deciding and causative are the newest functionalities that may alter present computing and networking. Therefore, sensors and RFID, among alternative technologies, are progressively deployed and can so permit integration of the real-world setting within the networked services. IoT can connect heterogeneous devices and can be dense, connecting billions of objects. An Internet-, IP- (Internet protocol) or TCP/IP (transport management protocol/Internet protocol) -based model stands in the middle of the IoT.

II. RFID TECHNOLOGY

Radio Frequency Identification system (RFID) is associate automatic technology and aids machines or computers to spot objects, record information or manage individual targets through radio waves [1]. The RFID technology was initial appeared in 1945, as associate spying tool for the state, that retransmitted incident radio waves with audio data. Similarly, the IFF (Identification Friend or Foe) electrical device developed within the UK was habitually utilized by the allies in war II to spot craft as a friend or foe. A generally RFID system has consisted of tags (transmitters/ responders) and readers (transmitters/receivers) [2]. The tag may be a semiconductor unit connected with associate antenna, which might be connected to associate objects because of the symbol of the article. The RFID reader communicates with the RFID tag victimization radio waves. The most advantage of RFID technology is that the automatic identification and information capture that guarantees wholesale changes across a broad spectrum of business activities and aims to scale back the price of the already used systems like bar codes. Though RFID technology was discovered a few years ago, it's advanced and evolved solely throughout the last decade since the price has been the most limitation all told implementations. As foretold [3], RFID is one of the massive opportunities in data technology, which can modify the planet broadly speaking and deeply. Once the RFID readers abided by applicable communication protocols area unit connected to the terminal of the web, the readers distributed throughout the planet will determine, track and monitor the objects connected with tags globally, mechanically, and in the period, if needed. This can be the alleged web of Things (IoT). The IoT refers to unambiguously classifiable objects (things) associated with their virtual representations in an Internet-like structure. The IoT initially became widespread through the Auto-ID Center and connected market analyst's publications. RFID is commonly seen as a necessity for the IoT. If all objects of the standard of living were equipped with radio tags, they may be known and inventoried by computers.

III. IoT AND RFID TECHNOLOGY

Internet of Things Internet of Things (IoT) is a global network infrastructure, linking physical and virtual objects through the exploitation of data capture and communication capabilities. It will offer specific object identification, sensor, and connection capability as the basis for the development of independent cooperative services and applications.

These will be characterized by a high degree of autonomous data capture, event transfer, network connectivity, and interoperability.

The IoT system architecture is generally divided into three layers: the perception layer, the network layer, and the service layer (or application layer), as shown in Figure.1

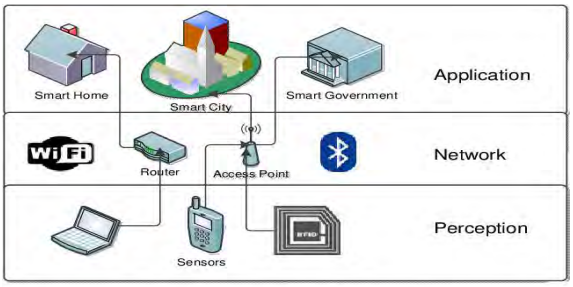


Figure 1. Three-layer IoT architecture.

Perception layer: It is the information origin and the core layer of IoT. All kinds of information of the physical world used in IoT are perceived and collected in this layer, by the technologies of sensors, wireless sensors network (WSN), tags and reader-writers, RFID system, camera, global position system (GPS), intelligent terminals, electronic data interface (EDI), objects, and so like.

Network layer: This layer, also called the transport layer, including access network and core network, provides transparent data transmission capability. By the existing mobile communication network, radio access network, wireless sensor network (WSN) and other communications equipment, such as global system for mobile communications (GSM), general packet radio service (GPRS), worldwide interoperability for microwave access (WiMax), wireless fidelity (WiFi), Ethernet, etc., the information form perception layer can be sent to the upper layer. At the same time, this layer provides an efficient, reliable, trusted network infrastructure platform to the upper layer and large scale industrial application [4].

Service layer: This layer, also called the application layer, includes the data management sub-layer and application service sub-layer. The data management sub-layer provides processing complex data and uncertain information, such as restructuring, cleaning and combining, and provides directory service, market to market (M2M) service, Quality of Service (QoS), facility management, geomatics, etc. by service-oriented architecture (SOA), cloud computing technologies, and so on. The application service sub-layer transforms information into content and provides a good user interface for upper-level enterprise applications and end-users, such as logistics and supply, disaster warning, environmental monitoring, agricultural management, production management, and so forth.

IV. APPLICATIONS OF RFID TECHNOLOGY

The functions of the RFID system generally include three aspects: monitoring, tracking, and supervising. Monitoring generally means to be aware of the state of a system, by repeated observing the particular conditions, especially to detect them and give warning of the change. Tracking is the observation of persons or objects on the move and supplying a timely ordered sequence of respective location data to a model. Supervising is the monitoring of the behaviors, activities, or other changing information, usually of people. It is sometimes done in a secret or inconspicuous manner. The RFID applications are numerous and far-reaching. The most interesting and successful applications include those for supply chain management, production process control, and object tracking management. Now RFID has been gradually and broadly used in the following fields.

Logistics and Supply
Manufacturing
Agriculture Management
Health Care and Medicine
Marine Terminal Operation
Military and Defense
Payment Transactions
Environment Monitor and Disaster Warning
Transportation and Retailing
Warehousing and Distribution Systems
Other applications in many walks of life businesses

For instance, in Manufacturing, RFID technology offers several applications in the automotive industry. An RFID based antitheft vehicle immobilizer is a protective device installed in many cars. RFID also holds great promise for the assembly and manufacturing processes of automobiles, in particular, for flexible and agile production planning, spare parts, and inventory management. RFID technology not only helps to automate the whole assembly process in which a significant reduction in cost and shrinkage can be achieved, but it also offers improved services to automobile users that include more efficient replacement part ordering and automated generation of maintenance reminders. The benefits that RFID offers to the automotive industry, both to the production process as well as to end-users, are visibility, traceability, flexibility, and added security.

V. CHALLENGES OF RFID TECHNOLOGY

Although promising, RFID is not without its challenges, which arise from both a technological and usage point of view. The functions of the RFID system usually embody 3 aspects: observance, tracking, and management. observance usually means that to remember of the state of a system, by continual observant the actual conditions, particularly to notice them and provides warning of the amendment, following is that the observation of persons or objects on the

move and supply a timely ordered sequence of various location information to a model. Management is that the observance of the behaviors, activities, or alternative ever-changing data, typically of individuals. It's generally worn out a secret or invisible manner. The RFID applications square measure various and extensive. The foremost fascinating and productive applications embody those for provide chain management, production method management, and object following management. Currently, RFID has been rising by step and broadly speaking employed in the subsequent fields.

Production
Agriculture M
Health Care and medication
Marine Terminal Operation
Military and Defense
Payment Transactions
surroundings Monitor and Disaster Warning
Transportation and marketing
deposit and Distribution Systems
alternative applications in several walks of life businesses

For instance, in producing, RFID technology offers many applications within the automotive business. RFID based anti-theft vehicle immobilizer protecting device can be installed in cars. RFID conjointly holds nice promise for the assembly and producing processes of vehicles, specifically, for versatile and agile production coming up with, spare components, and inventory management. RFID technology not solely helps to change the entire assembly method within which a major reduction in price and shrinkage will be achieved, however, it conjointly offers improved services to automobile users that embody additional economical replacement half ordering and automatic generation of maintenance reminders. The advantages that RFID offers to the automotive business, each to the assembly method furthermore on end-users, square measure visibility, traceability, flexibility, and other security.

A. Collision Problems

Communication between tags and readers area unit inherently prone to magnetism interference, coincidental transmissions in RFID result in collisions as readers and tags usually treat identical wireless channel. Therefore, economical anti-collision protocols for distinctive multi-tags at the same time area unit of nice importance for the event of large-scale RFID applications [1]. Many anti-collision protocols for RFID tag identification are projected, like question tree protocol (QT), binary tree protocol (BT), frame slotted ciao protocol (FSA), etc., however, most legendary protocols exhibit associate overall identification potency diminutive than five hundredths. Besides, uniform ID distribution has continually been assumed within the past. What is more, it helps remark the simplest performing arts options of RFID tag identification protocols, and for planning new and higher protocols. In [1], we tend to gift a unique and economical anti-collision protocol for RFID tag identification, i.e., collision tree protocol (CT), that outperforms all the opposite anti-collision protocols projected up to now.

B. Security and Privacy issues

Security and privacy problems [5] of RFID tags will affect each organization and people. Unprotected tags could also be at risk of eavesdropping, traffic analysis, spoofing or denial of service and plenty of additional. Even unauthorized readers will affect privacy by accessing tags while not enough access management, albeit the tag content is secure then additionally it is often caterpillar-tracked by the certain tag responses; "location privacy" is often stricken by a traffic analysis attack. the associate aggressor may also threaten the safety of systems, which depends on RFID technology through the denial of service attack. Due to its value and resource constraint limitations, RFID several researchers and scientists work to implement affordable security and privacy protocol to extend the relevance. Countless light-weight solutions are projected for RFID, however, they're still dear and at risk of security and don't resolve the safety problems. Thus there's an honest analysis scope within the field of planning associate economical ultra-lightweight scientific discipline protocol for affordable RFID systems.

VI. CONCLUSION

The IoT uses a spread of data sensing identification device and knowledge process instrumentation, like RFID, WSN, GPRS, etc. combining with the net to make an intensive network to info and intelligence the entities or objects. This paper analyzes the applications and challenges of RFID technology, which is the necessary and foundational element of IoT.

REFERENCES

[1] Xiaolin Jia ; Quanyuan Feng ; Taihua Fan ; Quanshui Lei, “RFID technology and its applications in Internet of Things (IoT),” 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet),IEEE

[2] X.L. Jia, Q.Y. Feng, C.Z. Ma, “An efficient anti-collision protocol for RFID tag identification,” IEEE Communications Letters, vol.14, no.11, pp.1014-1016, 2010.

[3] K. Finkenzeller, RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification, and Near-Field Communication, New York: Wiley, 2010.

[4] D.Culler, “10 Emerging Technologies That Will Change the World,” Technology Review, pp.33-49, Feb. 2003.

[5] Y.X.Lu, T.B.Chen, Y.Meng, "Evaluation guiding system and intelligent evaluation process on the Internet of Things,” American Journal of Engineering and Technology Research, vol.11, no.9, pp.537-541, 2011.

[6] R.K.Pateriya, S.Sharma, “The Evolution of RFID Security and Privacy:

A Research Survey,” 2011 International Conference on Communication

Systems and Network Technologies (CSNT), pp.115-119, June 2011

Responsible and Scalable Approach for Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud: A Survey

Swati Abhang
Thakur College of Engineering & Technology, Mumbai, India
swatiabhang@gmail.com.

Abstract: **In Today's world Cloud Computing which moves the application software and databases to the centralized large data centers, where the management of the data and services may not be fully trustworthy. Several trends are opening up the era of Cloud Computing, which is an Internet-based development and use of computer technology. One of the biggest concerns with cloud data storage is that of data integrity verification at untrusted servers. To preserve data privacy, a basic solution is to encrypt data files, and then upload the encrypted data into the cloud. To resolve this problem recently the best efficient method MONA presented for secured multi owner data sharing in However we identified some limitations in the same approach in terms of reliability and scalability. Hence in this paper we are further extending the basic MONA by adding the reliability as well as improving the scalability by increasing the number of group managers dynamically.**

Keywords– Cloud Computing, dynamic groups, data sharing, reliability, integrity, scalability.

I. INTRODUCTION

A) What is cloud computing?

Cloud computing is one of the greatest platform which provides storage of data in very lower cost and available for all time over the internet Cloud computing is Internet-based computing, whereby shared resources, software and information are provided to computers and devices on demand. Several trends are opening up the era of Cloud Computing, which is an Internet-based development and use of computer technology. Cloud computing means more than simply saving on IT implementation costs. Cloud offers enormous opportunity for new innovation, and even disruption of entire industries. Cloud computing is the long dreamed vision of computing as a utility, where data owners can remotely store their data in the cloud to enjoy on-demand high-quality applications and services from a shared pool of configurable computing resources.

B) Basic Concept

Maintaining the integrity of data plays a vital role in the establishment of trust between data subject and service provider. Although envisioned as a promising service platform for the Internet, the new data storage paradigm in "Cloud" brings about many challenging design issues which have profound influence on the security and performance of the overall system. One of the biggest concerns with cloud data storage is that of data integrity verification at untrusted servers. What is more serious is that for saving money and storage space the service provider might neglect to keep or deliberately delete rarely accessed data files which belong to an ordinary client. Consider the large size of the outsourced electronic data and the client's constrained resource capability, the core of the problem can be generalized as how can the client find an efficient way to perform periodical integrity verifications without the local copy of data files. To preserve data privacy, a basic solution is to encrypt data files, and then upload the encrypted data into the cloud [2]. CS2 provides security against the cloud provider, clients are still able not only to efficiently access their data through a search interface but also to add and delete files securely. Several security schemes for data sharing on untrusted servers have been proposed secure file system designed to be layered over insecure network and P2P file systems such as NFS, CIFS, Ocean Store, and Yahoo! Briefcase.

C) Advantages and Disadvantages of Cloud Computing:

Advantages: -
24/7 Support
Easy to Maintain.
Secure Storage and Management
Location Independent
Less cost (Pay-as-per-you-Use).
high level computing
Personalized Backup and recovery.
Remote access.
Green computing.
Disadvantages: -
Lack of control
Security and privacy.

b) S. Kamara and K. Lauter [3] in this paper consider the problem of building a secure cloud storage service on top of a public cloud infrastructure where the service provider is not completely trusted by the customer. They describe, at a high level, several architectures that combine recent and non-standard cryptographic primitives in order to achieve our goal. Survey the benefits such architecture would provide to both customers and service providers and give an

overview of recent advances in cryptography motivated Higher operational cost.Reliability

II. LITERATURE SURVEY

In the literature survey we are going to discuss some existing technique for cloud.

a) M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia [2] the data centers hardware and software is what we will call a cloud. When a cloud is made available in a pay-as-you-go manner to the general public, they call it a public cloud; the service being sold is utility computing. They use the term private cloud to refer to internal data centers of a business or other organization, not made available to the general public, when they are large enough to benefit from the advantages of cloud computing that we discuss here. Thus, cloud computing is the sum of SaaS and utility computing, but does not include small or medium-sized data centers, even if these rely on virtualization for management. People can be users or providers of SaaS, or users or providers of utility computing. They focus on SaaS providers (cloud users) and cloud providers, which have received less attention than SaaS users.attribute-based encryption (ABE), proxy re-encryption, salient properties of user access privilege confidentiality and user secret key accountability.

d) E. Goh, H. Shacham, N. Modadugu, and D. Boneh [5] the use of SiRiUS is compelling in situations where users have no control over the file server (such as Yahoo! Briefcase or the P2P file storage provided by Farsite). They believe that SiRiUS is the most that can be done to secure an existing network file system without changing the file server or file system protocol. Key management and revocation is simple with minimal out-of-band communication. File system freshness guarantees are supported by SiRiUS using hash tree constructions. SiRiUS contains a novel method of performing file random access in a cryptographic file system without the use of a block server. Extensions to SiRiUS include large scale group sharing using the NNL key revocation construction. e)R. Lu, X. Lin, X. Liang, and X. Shen [6] in this paper secure provenance is of paramount importance to the flourish of cloud computing, yet it is still challenging today. In this paper, They formally defined the secure provenance and the corresponding security model in cloud computing. Then, in proposed a concrete secure provenance SP scheme based on the bilinear pairings, and used the provable security technique to prove its security in the standard model. Due to its comprehensive security features, the proposed SP scheme provides trusted evidences for data forensics in cloud computing and thus pushes the cloud computing for wide acceptance to the public.

f) B. Waters [7] presented the first cipher text-policy attribute-based encryption systems that are efficient, expressive, and provably secure under concrete assumptions. All of our constructions fall under a common methodology of embedding an LSSS challenge matrix directly into the public parameters. Our constructions provide a trade off in terms of efficiency and the complexity of assumptions.

g) V. Goyal, O. Pandey, A. Sahai, and B. Waters [8] they develop a new cryptosystem for One-grained sharing of encrypted data that call Key-Policy Attribute-Based Encryption (KP-ABE). In cryptosystem, cipher texts are labelled with sets of attributes and private keys are associated with access structures that control which cipher texts a user is able to decrypt. They demonstrate the applicability of our construction to sharing of audit-log information and broadcast encryption. Our construction supports delegation of private keys which subsumes Hierarchical Identity-Based Encryption (HIBE).

h) A. Fiat and M. Naor [9] they introduce new theoretical measures for the qualitative and quantitative assessment of encryption schemes designed for broadcast transmissions. The goal is to allow a central broadcast site to broadcast secure transmissions to an arbitrary set of recipients while minimizing key management related transmissions. They present several schemes that allow centers to broadcast a secret to any subset of privileged users out of a universe of size so that coalitions of users not in the privileged set cannot learn the secret.

i)B. Wang, B. Li, and H. Li, [10] in this paper, we propose Knox, a privacy-preserving auditing scheme for shared data with large groups in the cloud. They utilize group signatures to compute verification information on shared data, so that the TPA is able to audit the correctness of shared data, but cannot reveal the identity o f the signer on each block. With the group manager's private key, the original user can efficiently add new users to the group and disclose the identities of signers on all blocks. The efficiency of Knox is not affected by the number of users in the group.

j) D. Pointcheval and J. Stern [11] As Explained in the Introduction, there w ere several proposals for pro v ably secure Signature schemes. However, in all cases, the security was at the cost of a considerable loss in terms of efficiency. Concerning blind signatures, Damgard, Ptzmann and Waidner and more recently at Crypto '97, Juels et al. Have presented some blind signature schemes with a complexity-based of security. Again, the security y is at the cost of inefficiency. In the weaker setting by the random oracle model, we have provided security arguments for practical and even efficient digital signature schemes and blind signature schemes. On the ground of our reductions, one can justify realistic parameters, even if they are not optimal. Further improvements are expected particularly in the case of blind signatures where it should be possible to

obtain a reduction polynomial in the size of the keys and in the number of interactions with the signer

III. EXISTING SYSTEM

In the literature study we have seen many methods for secure data sharing in cloud computing, however most methods failed to achieve the efficient as well as secure method for data sharing for groups. To provide the best solutions for the problems imposed by existing methods, recently the new method was presented called MONA [1]. This approach presents the design of secure data sharing scheme, Mona, for dynamic groups in an untrusted cloud. In Mona, a user is able to share data with others in the group without revealing identity privacy to the cloud. Additionally, Mona supports efficient user revocation and new user joining. More specially, efficient user revocation can be achieved through a public revocation list without updating the private keys of the remaining users, and new users can directly decrypt files stored in the cloud before their participation. Moreover, the storage overhead and the encryption computation cost are constant. Therefore practically in all cases MONA outperforms the existing methods.

Fig 3.1 Existing System Model

3.1 Disadvantages of Existing System

However as per reliability and scalability concern this method needs to be workout further as if the group manager stop working due to large number of requests coming from different groups of owners, then entire security system of MONA failed down.

IV. PROPOSED SOLUTION

Thus to achieve the reliable and scalable MONA approach, in this project we are presenting the new framework for MONA called as RS-MONA (Reliable Scalable-MONA). In this method we are further presenting how we are managing the risks like failure of group manager by increasing the number of backup group manager, hanging of group manager in case number of requests more by sharing the workload in multiple group managers. This method claims required efficiency, scalability and most importantly reliability.

Fig 4.1 Proposed System Model

4.1Advantage of Proposed System

To overcome the disadvantage of existing system MONA, the proposed RS-MONA is if the group manager stop working due to large number of requests coming from different groups of owners, then backup group manager will remains available.

V. SYSTEM REQUIREMENT & SPECIFICATION

5.1 Software Requirements:
Front End: Java
Tools Used: Eclipse/Net beans
Operating System: Windows 7
5.2 Hardware Requirements:
Processor: Pentium IV 2.6 GHz
Ram: 512 Mb
Monitor: 15" Colour Hard Disk: 20 Gb Floppy Drive: 1.44 Mb Keyboard: Standard 102 Keys Mouse: 3 Button

VI. CONCLUSIONS

In conclusion, cloud computing is very attractive environment for business world in term of providing required services in a very cost effective way. However, assuring and enhancing security and privacy practices will attract more enterprises to world of the cloud computing In Thus to achieve the reliable and scalable MONA approach; in this paper we are presenting the new framework for MONA called as RS-MONA (Reliable Scalable-MONA). In this method we are further presenting how we are managing the risks like failure of group manager by increasing the number of backup group manager, hanging of group manager in case number of requests more by sharing the workload in multiple group managers. This method claims required efficiency, scalability and most importantly reliability. Extensive analyses show that our proposed scheme satisfies the desired security requirements and guarantees efficiency as well.

References

[1]. Xuefeng Liu, Yuqing Zhang, Boyang Wang, and Jingbo Yan, "Mona: Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO. 6, JUNE 2013.

[2]. M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz,A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, Apr. 2010.

[3]. S. Kamara and K. Lauter, “Cryptographic Cloud Storage,” Proc. Int’l Conf. Financial Cryptography and Data Security (FC), pp. 136- 149, Jan. 2010.
[4]. S. Yu, C. Wang, K. Ren, and W. Lou, “Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing,” Proc. IEEE INFOCOM, pp. 534-542, 2010.
[5]. E. Goh, H. Shacham, N. Modadugu, and D. Boneh, “Sirius: Securing Remote Untrusted Storage,” Proc. Network and Distributed Systems Security Symp. (NDSS), pp. 131-
[6]. 145, 2003.
[7]. R. Lu, X. Lin, X. Liang, and X. Shen, “Secure Provenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing,” Proc. ACM Symp. Information, Computer and Comm. Security, pp. 282-292, 2010
[8]. B. Waters, “Cipher text-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization,” Proc. Int’l Conf. Practice and Theory in Public Key Cryptography Conf. Public Key Cryptography, http://eprint.iacr.org/2008/290.pdf, 2008.
[9]. V. Goyal, O. Pandey, A. Sahai, and B. Waters, “Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data,” Proc. ACM Conf. Computer and Comm. Security (CCS), pp. 89-98, 2006.
[10]. Fiat and M. Naor, “Broadcast Encryption,” Proc. Int’l Cryptology Conf. Advances in Cryptology (CRYPTO), pp.480-491, 1993.
[11]. Wang, B. Li, and H. Li, “Knox: Privacy-Preserving Auditing for Shared Data with Large Groups in the Cloud,” Proc. 10th Int’l Conf. Applied Cryptography and Network Security, pp. 507-525, 2012.
[12]. D. Pointcheval and J. Stern, “Security Arguments forDigital Signatures and Blind Signatures,” J. Cryptology, vol.13, no. 3, pp. 361-396, 2000.

Error Control in Data Link Layer-A Roadmap

Chitvan Mehrotra, Sowmyashree
Thakur college of engineering and technology, Mumbai, India
chitvanmehrotra@gmail.com, sshree87@gmail.com

Abstract—**In today's world, accuracy is the most important key of any data transfer on a computer network. We access a network on this parameter only and we expect that the receiver receives the same data as sent by the sender. If the data is changed while sending or is hampered when the data is sent from one station to the other we would not accept that computer network as a reliable one. In this paper we would highlight the datalink layer of the OSI model and would discuss in detail the error control mechanisms in it. When data is sent from one node to the other how do we detect errors and how do we correct it in case the data has been hampered.**

Keywords – Error control; Error detection; Error correction; Data link layer error control mechanisms.

I. INTRODUCTION

Data link layer has two main subparts or sublayers having different functionalities, they are namely logical link control (LLC) and media access control (MAC)

Logical Link Control: It deals with the various protocols of the network, flow-control mechanisms, and error control techniques on the network. It tries to establish the best possible link between the users or stations [1]

Media Access Control: It deals with actual control of media, how the data has to be sent on the network and in case of any clashes on the network which station would get access on the transmission media is decided by this sublayer of the data link layer.

Data link layer handles very complex functions on our computer network. First of all it works for abstraction and hides the complex details of the hardware from other layers, various other functions served by the Data link layer are as follows:-

- Framing

The data link layer receives the packets of the above network layer and changes them into frame structure, after the frames are made this layer has the responsibility of passing these packets to the hardware which lies beneath in the system.

- Synchronization

When the frames made by this layer are sent over in the computer network it has the responsibility to make sure that the receiver and sender are both synchronised.

- Error Control

When we are sending data there are chances that some bits of the data may change during the transmission from one station to the other because there are various errors which can be introduced in the system. So the data link layer has the responsibility to try identifying these errors which have occurred and try correcting them as well [2] [3]
Flow Control

The speed of the data sent over the channel should be same as the speed the receiver receives because if there will be difference in the speeds there can be problems in the system and congestion can also take place.

- Multi-Access

When a single channel is being used by multiple users in the system then various multiple access protocols can be used which are monitored by this layer. CSMA/CD is used for this purpose.

In this paper we would discuss about the error control functionality of the data link layer in detail. So to understand error control we first need to understand what kind of errors can be introduced when we are sending data from one end to the other Types of Errors [4]

There may be three types of errors:

- Single bit error

If only single or one bit of the frame of data is hampered or changed then it is a single bit error**.**

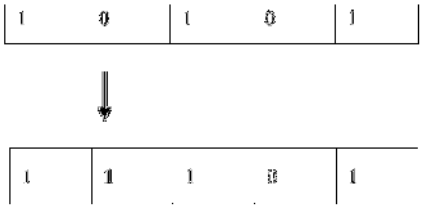


Figure 1: single bit error in the data stream

- Burst error

If there is error in more than one bits in the data being sent then it is called burst error where the

length of the burst is from the first bit where the data got corrupted to the last bit.

If we have such corrupt data being received by the received we have to apply certain mechanism to detect these errors and finally correct them [5]

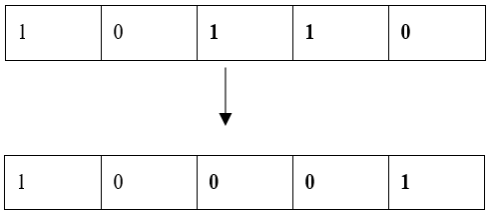


Figure 2: burst error in a data stream

II. ERROR DETECTION

Error control mechanism may involve two possible ways:

- Error detection
- Error correction

There are three most common types of error detection techniques or schemes:-parity check, checksum and cyclic redundancy check (CRC).We would now discuss them one by one

a) Parity check

Parity of a string can be determined by the number of 1's present in any data string or rather binary string which consists of 0's and 1's.The total number of 1's determine the parity of that string.[6]

We add a bit known as the parity bit or it is sometimes also called a check bit to make sure that the total count of the 1's is either even or odd number.Parity bits is the simplest way to detect error although it cannot be used to correct error it can only be used to detect it.Let us take one example:-

if a system checks the data 0111001 for even parity, the parity bit should be 0 if the data is correct because in this case the number of 1's is already even.If we had to produce the same data with odd parity then the parity bit would be 1 to make the number of 1's odd in the string

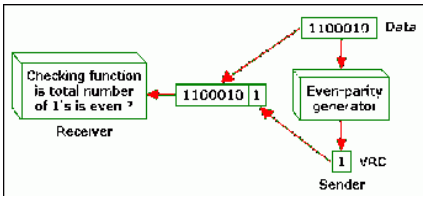


Figure 3: Checking even parity in the data

Here the data at the sender side was 110010 and we are using even parity as the number of 1's in the above string are 3 to make it even we will add 1 to the data and transmit it to the other end. [7]. The receiver will also

Original Data	Even Parity	Odd Parity
00000000	0	1
01011011	1	0
01010101	0	1
11111111	0	1
10000000	1	0
01001001	1	0

Figure 4: Parity checker

b) Checksum

- In this scheme we first divide the data string into K segments each of them having M number of bits.
- The next step is to add these segments together using the 1's compliment method, in 1's compliment method if a carry is generated by the addition then it is also added with the sum. So like this we will calculate the sum of all the frames or segments made in step 1.
- Now the sum is complemented and is then sent to the receiver.[8][9]
- At the receiving end again all the data frames are added along with the checksum value. Now this
- Final sum is complemented, if we get all zeroes we say that no error has occurred during transmission of data and the data is accepted. If we do not get all zeroes the data sent by the sender is discarded.

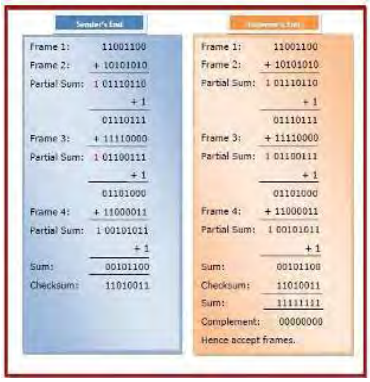


Figure 5: Checksum at sender and receiver

c) CRC

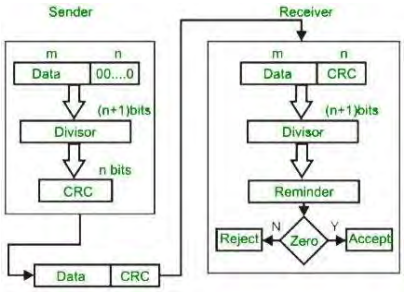


Figure 6: CRC procedure

Another error control technique is cyclic redundancy check in which we mainly perform modulo-2 binary division.
Steps of CRC are as follows:-

Step1-first of all we have to mutually decide a divisor, the sender and the receiver agree to a common divisor which is shared among the parties involved.
Step 2- add (n-1) zeroes at the end of the dividend
Step 3-the data to be sent appended with the the zeroes forms the final dividend
Step 4- Perform the modulo 2 division which is rather an XOR operation and finally we get the result as the remainder. So the remainder of this division is called the CRC. It is added to the data in the end and the whole block is sent to the receiving side.
Step-5 At the receiving end again a modulo 2 division is performed with the data and the divisor predecided by the sender and the receiver.
Step-6 The final step of this entire process is to verify whether the data received was correct or not. If after division we get all zeroes the data received is correct and therefore the data was not hampered during transference. On the other side of this if we do not get all zeroes and we get some other remainder at the receiver end we can easily say that the data was changed during the transmission from one end to the other and it is hampered. But we cannot correct the data from this method we can just tell that the data received is correct or not .[10]

III. CONCLUSION

In this survey paper we have discussed the various error detection techniques in detail. By error detection we actually mean that the data sent from the sender side to the receiving side is correct or not.
[11] As we all know due to noise in the channel or other interferences the data can be hampered when it is being sent over the network. Even the safest of media cannot guarantee 100%error free transfer, because of this reason the role of error detection and correction techniques is important. We have tried to analyse these techniques to understand that even if the data is changed over the network how can we actually detect that the error has occurred.

REFERENCES

[1] Pramod S P, Rajagopal A, Akshay S Kotain,'FPGA Implementation of Single Bit Error Correction using CRC', International Journal of Computer Applications (0975 – 8887) Volume 52– No.10, August2012
[2] Nidhi Syal, Vaneet Chahal, 'Comparison of "reed_muller and hamming codes" for error detection and correction', ISP Journal of Electronics Engineering, Vol1, Issue 1, October'2011, (5-9)
[3] Md. Sharif Uddin, Cheol Hong Kim, Jong- MyonKim, 'Accelerating SoftDecision Reed-Muller Decoding Using a Graphics Processing Unit', Asiapacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology Vol.4, No.2, December (2014), pp. 369-378
[4] Varsha P. Patil, Prof. D. G. Chougule, Radhika.R.Naik, 'Viterbi Algorithm for error detection and correction', IOSR Journal of Electronicsl and Communication Engineering (IOSR-JECE) ISSN: 2278- 2834-, ISBN: 2278-8735, PP: 60-65
[5] Notes on 'Error Detection and Correction ', Version 2 CSE IIT, Kharagpur
[6] Khalid Faraj, 'Design Error Detection and Correction System based on Reed_Muller Matrix for Memory Protection', International Journal of Computer Applications (0975 – 8887) Volume 34– No.8,
November 2011
[7] Othman O. Khalifa, Aisha-Huassan Abdullah, N. Suriyana, Saidah Zawanah and Shihab A. Hameed 'Reed- Muller Codec Simulation Performance', Journal of Computer Science 4 (10): 792-798, 2008 ISSN 1549- 3636
[8] Reed, I. S., 'Class of multiple error correcting codes and their decoding scheme', Institute of Radio Engineers Transaction on Information Theory, PGIT-4: pp. 38-49, 1954.
[9] Muller, D. E., 'Application of Boolean algebra to switching circuit design and to error detection', Institute of Radio Engineers Transaction on Electronic Computers, EC-3: pp. 6-12, September 1954. Faraj, K., Almaini, A.E.A.,'Minimization of Dual Reed-Muller Forms using Dual Property' WSEAS TRANSACTIONS on CIRCUITS and SYSTEMS, Issue 1, Vol. 6, pp.9-15, January 2007.
[10] Faraj, K., Almaini, A.E.A., 'Optimal Expression for Fixed Polarity Dual Reed-Muller Forms', WSEAS TRANSACTIONS on CIRCUITS and SYSTEMS, Issue 3, Vol. 6, pp.9-15, March 2007.

Framework for Blockchain based Land Registration System.

Sandip Bankar, Neerajkumari Khairwal
Thakur College of Engineering and Technology, Kandivali

Abstract – **Land registration leads to registration of all land transac- tions/deeds undertaken in any geographical jurisdiction of a governing authority. In most countries, these records are maintained with various government and local bodies which at times leads to title disputes and costly litigation. The current land registry system is widespread with corruption and inefficiencies.**

There are many citizens who don't have confidence in the system. There are some situations, where the disaster like fire, flood, and earth- quake destroyed paper records. While many countries and organizations did their best to help to rebuild the nation, there were thousands of plots of land where the rightful owner could not be identified, and in many instances ownerships were in dispute. Hence, there is need of real time record maintenance.

Keywords: Land Records, Blockchain, Revenue Departments.

I. INTRODUCTION

Land is an important resource as its one of a unique due to its location, area, and with increasing population, its interest continues to expand, while its supply is restricted. Access to land (or land rights) has a wide-going effect on jobs, modern, financial, and social development. It has been noticed that individuals with broad rights to arrive are in an ideal situation than the landless, because of better access to business sectors and other financial open doors that accompany arrive rights.

Land ownership is comprehensively characterized by the entrance to a land title. Land title is an archive that decides the responsibility for a resolute property. Having an unmistakable land title secures the privileges of the title holder against different cases made by any other individual to the property. In India, arrive proprietorship is re- solved through different records, for example, deal deeds that are enrolled, property charge papers, government overview records, and so on. [1]

Be that as it may, land record titles in India are indistinct because of different reasons, for example, heritage issues of the zamindari framework, holes in the legitimate struc-ture, and poor organization of land records. This has prompted a few legitimate ques- tion identified with land proprietorship, and influenced the farming and land areas. Such issues have additionally featured the significance of having clear land titles, and an efficient land records system

Land registration for the most part describes matters concerning proprietorship, own- ership or different rights in land can be recorded (normally with an government office or division) to give proof of title, encourage exchanges and to prevent unlawful trans- fer.

Land Administration as characterized as is the way toward deciding, recording and maintaining data about possession, cost and utilization of land, when executing land registration strategies. This incorporates procedures of land enrollment, valuation and the land register. Governments seek to enhance benefits in Land Administration keep on presenting innovation and new advancements in techniques of land registration systems. Different governments have started using various technologies for easy and secure land registration systems [2].

Land registration system becomes successful when there is equal trust in all partners involved in the system. In some countries people have less trust in current land regis- tration system. There are some cases where frauds and corruptions happening in cur- rent land registration systems. [4]

1.1 The requirement of clear land titles

1. *High prosecution*: A World Bank analysis from 2007 states that total land related cases represent 66% of all pending court cases in the country.[5] These land prob-lems incorporate those identified with the legitimacy of land titles and records, and legitimate proprietorship. A NITI Aayog paper recommends that arrive debate by and large take around 20 years to be resolved.[6] Land related cases add to the weight of the courts, tie up land in prosecution, and further effect divisions and tasks that are reliant on these debated land titles.

2. *Agricultural importance*: Land is regularly utilized as guarantee for acquiring loans by people. It has been seen that debated or doubtful land titles restrain supply of capi-tal and credit for agriculture.[7] Small and peripheral farmers, who represent the greater part of the all-out land possessions, and may not hold formal land titles, can't get to systematized credit.[8]

1.2 The Need of Real Time Record Maintenance System

Land records are important from a perspective of land title with one owner, trail of all transactions done regarding the unit and trustworthy system to provide shares of own- ers of units.

Trustworthy system: Many citizens simply don't have confidence in the system. Some are unsure if they legally own a piece of land, even if they have a legitimate sale deed. Others who want to buy a piece of land are not sure if the seller legally owns it.

Fast process: Current system takes one or two months to complete the transaction. There is need of system which could complete transaction in less time.

Decentralized rights: There should be decentralized rights of the sale transac- tions/deeds.

Risk of Multi-owner units: There have been instances where an owner sells same property to multiple buyers at same time, hence, leading to ownership dispute

1.3 Current Challenges:

Data Integration: Currently government officials have data records of only respective zones. Moreover, data formats and procedures differ across various zones. There's no uniformity across country for land records.

Ownership according to share: In case of property having multiple legitimate owners, it is a challenging task to maintain the information that how much share of property is owned by each individual.
Scalability: Current land record maintenance system is not scalable enough to main- tain record of entire country.
Security: Land Records need to be securely maintained for ownership integrity as they
are widely used by other departments such as land acquisition departments, banks and many more.
Accessible to different departments: Since land records are used by various depart- ments mentioned above, authorized departments need to have easy access and appro- priate rights to these land records.
Automatic Ownership/ Checking: Current land record maintenance system lacks the functionality to validate ownership of property or user wise owned properties.
User Friendly Access: Users are currently dependent on government officials and/or agents to retrieve land ownership records. Hence there is need for easy access to users [10].

II. TRADITIONAL LAND RECORD MAINTENANCE SYSTEM

Traditional land registration system consist of following steps as shown in figure
1,where physically user needs to go to the offices to carry out pro-cess.

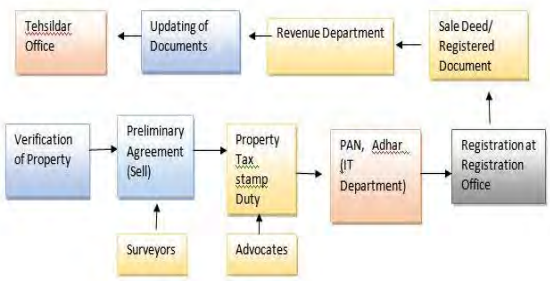


Fig. 1. Traditional land Registry system

2.1 Procedure for registration of land property in India

Registration process in India has become simple as the introduction of computerized system are used. This computerized system can help users to eliminate the middleman and it brings transparency in execution

However introduction of computerized system it has not completely replaced the manual paperwork and in some states it is necessary to submit the application manual- ly to the respective department or authority. The users need to download application form from the website of the concerned department or the concerned office.
The registration process comprises of paying registration fee for sale deed and it is necessary to record all the documents legally with the sub registrar office.it varies based on property is purchased directly from developer or secondary sale. Because in the other case process involves stamp duty and registration of transfer deed.

2.2 Steps:

Verification of the property:
Buyer needs to verify the property by physically visiting or any other option. It is based on whether property is purchased from secondary seller or developer.

Preliminary agreement:
Once property is verified then buyer and seller collectively prepare the preliminary agreement with the help of an advocate

Property tax stamp duty:
After completion of preliminary agreement it is required to pay stamp duty charges and registration fee. It varies from state to state. It may range from three to seven percent of property value. Property value can be decided from the annually ready rancor register.

Submission of documents:
Buyers and sellers with 2 witnesses need to submit their identity proofs and necessary documents.

Execution and registration of sale deed:
Next step involves execution of sale deed and register the sale deed. It needs to visit sub register office and require to submit the photographs, relevant documents like photograph of plot, 7/12 document or property card.

Property Mutation:
This is last step to complete registration process which involves getting mutation of property done in the land and revenue records. For this it is required to visit local municipal authorities.
This complete process may takes 20 to 30 days to complete

III. PROPOSED REAL TIME RECORD MAINTENANCE PROCESS:

Blockchain technology can be used to develop immutable distributed ledger. Every Department interacts with the blockchain using a public-private cryptographic key combination as shown in figure 2.

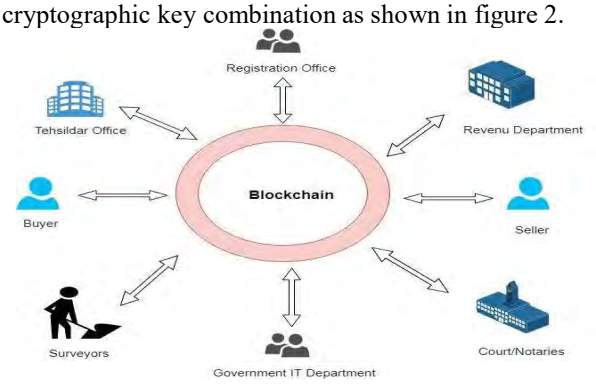


Fig. 2: Property Buying & Selling Process through Blockchain.

3.1 Process:

Step-1 Buyer searches a property on Blockchain platform.
Step-2 Buyer verifies the selected property.
Step-3 Buyer completes the payment to seller bank. Step-4 Registration will be done at office.
Step-5 Then seller initiates the blockchain transaction and broadcast the transac- tion.
Step -6 transaction will be verified by users on blockchain.
Step -7 block will be created and added to blockchain.

IV. PROPOSED SYSTEM

The proposed system, which we have proposed here incorporates many key benefits of blockchain technology, such as Immutable transaction records so no one can doubt the authenticity of transactions, records are permanently stored on the blockchain so no one can tamper these transaction records and there transaction records can be seen by anyone, anytime based on the privileges. Blockchain is a form of distributed ledger where data is stored on distributed network of computers. With blockchain anyone can access the network which automatically receive changes made to any database.

4.1 Smart contracts:

Smart contract is computer program which allows parties that are under certain condi- tions to directly control the transfer of assets or digital currencies.

These are self-executing programs used in blockchain to handle specific task. Actions that are kept in smart contracts are executed based on some events triggered.
They help you to facilitate, verifies and enforce negotiation or performance. Smart contracts are able to track performance in real life and can be used save the cost very effectively.
They can be used to:

1. To minimize the transaction cost.
2. Greater security.
3. To reduce overburden on trusted intermediaries.
4. To turn legal suggestions into automated processes.

Blockchain can be used in land registry to record all the transactions that happened in land record system. As blockchain is tamperproof and it cannot be modified by any illegitimate user as the system only consists of legitimate users who are allowed to transfer the property to other user.

Here smart contract can be used to automatically transfer the property when buyer deposits money into seller accounts.

ESCRO account:

This account can be used as intermediary account to temporarily hold the amount. It will then transfer amount to the seller account once the deed is done. If any stake-holder disallows the transaction then money will be deposited back to buyer account.

V. PROPOSED ARCHITECTURE

Our proposed system involves inclusion of departments like registration office, reve- nue department, court/notaries, surveyors, Tahsildar office as shown in above archi- tecture.

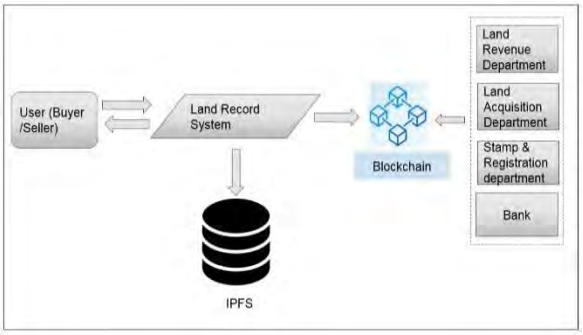


Fig 3: General Architecture of system

The proposed architecture consists of :

1. User: User is the one who is going to use the application in order to register his/her property or can a buyer or a seller of a property.
2. Land Record System: It is the graphical user interface which interacts with both the end user and the system kernel
3. IPFS: IPFS stands for Interplanetary File system is distributed system which is used for storing and accessing files, websites, applications and data. IPFS is used to store the scanned documents and id proofs in the system. It returns the cryptographic hash value of the file which is then passed to the block- chain to store it with the other data.
4. Blockchain: It is a chain of blocks which contains transactions which are linked to its previous node with the help of hash value of the previous block. The transactions of the related to lands will be stored in the blockchain.
5. Stakeholders (validators): Stakeholders are the validators of the transaction based on the authenticity of the documents the stakeholders will validate the transaction. The stakeholders include the Land Revenue department, Land Acquisition Department, Stamp and Registration Department and bank.

5.1 First time registration

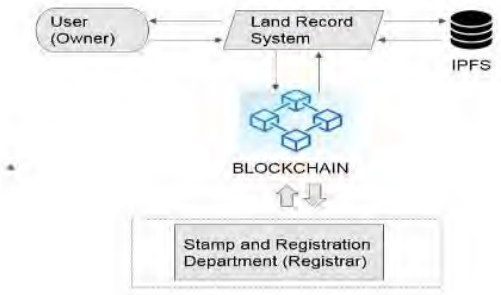


Fig 4: First Time Registration Architecture

Every user or the land owner need to register his property in the Land Record system for the first time irrespective of his interest in selling or buying any property.

First Time Registration is required in order to maintain the record of the land owner- ship. The stakeholder who is responsible for validating the transactions will be the registrar from the Stamp and Registration Department. Registrar will validate the transaction only after testing the authenticity of the documents which are uploaded by the User. Only when the registrar finds that the documents are authentic then and only then he will validate the transaction and block will be created and appended to the blockchain network.

Steps involved in the first time registration are:
Step 1- User will click on the create account and an account number will be generated Step 2- User should enter the personal details which includes name, address, contact details, government ID number etc. along with account that he has received in Step 1. Step 3- User then will enter the property related information which involves survey number, area, owner name, taluka, village, type of occupancy and documents of proof of ownership (property papers) along with document of ID proof will be uploaded by the User. Documents will be stored in the IPFS and cryptographic hash will be send to the LR system.
Step 4- User will click on the Validate button
Step 5- On clicking block will be created and transferred for validation to the regis- trar. Registrar will validate the property and Identification documents along with de- tails for its authenticity.

Step 6- If the documents are found to be authentic then validation status "true" will be displayed to the user and Block created will be broadcasted and appended to the blockchain. Else if documents are not authentic then "false" status will be displayed and the block will be dropped.
Step7- Stop.

5.2 Transfer of property

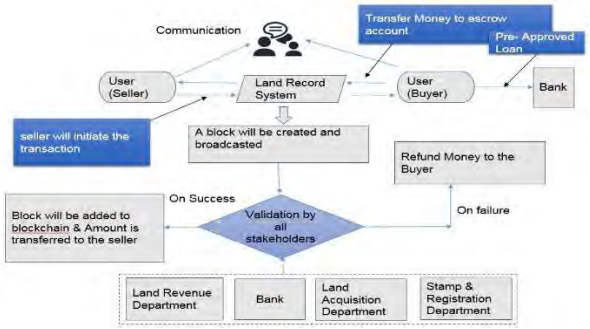


Fig 5: Architecture of Land transfer

The other type of the user is the buyer or the seller who are going to us the Land Rec- ord system for the transaction. In this case the seller and the buyer of the property has already communicated about the deal of the property. Here the seller is supposed to initiate the transaction and the buyer will have a pre-approved loan as shown in figure

5. Here all together there will be all 4 validators which include

1. Land Revenue Department: LRD will check whether the owner has not paid the property tax or does not have any previous dues left.
2. Bank: Bank will check whether the property is currently mortgage by the owner for some loan or not. Also it is responsible to validate that all the in- stallments related to mortgage are being paid by the current owner or not.
3. Land Acquisition Department: LAD will check if the property is the govern- ment property or is in the list of the Land to be acquainted or acquire by the government
4. Stamp & Registration Department: SRD will validate the Survey No and documents for selling the land.

5.3 Steps involved in selling the land

Step 0- The seller and the purchaser has previously communicated and Purchaser has a previously approved loan from the bank.
Step 1- The seller will initiate the transaction by providing the Account no. of the buyer on blockchain and block will created. The Buyer will transfer the Money to the Escrow account of the Registrar from the Stamp and Registration Depart- ment.
Step 2- The created block will be validated by the 4 stakeholders on parameters mentioned above.

Step 3-
1. After block is successfully validated and status is success then the transaction is successful. After which the amount from the Escrow account will be trans- ferred to the seller's account and property will be transferred on the name of its buyer. Then new block will be added to the blockchain.
2. In case the validation status is fail then the amount from the escrow account will be transferred back to the buyer and block will be dropped.

Step 4- Transaction Complete.

5.4 Technology Stack: Ethereum, Solidity, Remix.

VI. RESULTS AND DISCUSSION

We have implemented the proposed system using Ethereum environment with solidity language and stored records on IPFS as discussed in proposed system.Some screen- shots are shown below Figure 6 indicates login screen where first time registration of users is possible while Figure 7 indicates Ethereum accounts list .

Fig 6:Screenshot for First time registration form

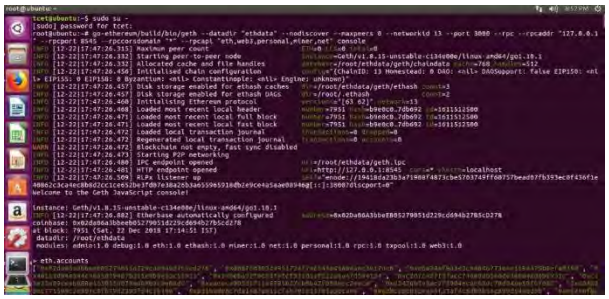


Fig 7: Screen shot for ethereum accounts

VII. CONCLUSION

Blockchain technology in land registry brings transparency in administration. Prove- nance of land can be maintained easily which can also help to find the real owner of the land and thus helping to solve the land ownership cases. Land records across country can be easily integrated and validated – For an owner having multiple proper ties in different regions integrating those data together will help to find the total num- ber of land assets under that owner data integration and overall validations can be done. Synchronization among various concerned departments can be achieved – In case of blockchain based system for land records various stakeholders i.e. various concerned departments acts as the validators during each process because of which the inconsistency that used to be present in the earlier system gets removed and thus synchronization in the departments will be achieved. Common man do not need to move to different offices – Earlier the common man had to move to different offices in order to transfer the property ownership and also had to hire an agent to make the process more faster. This thus leads to frustration to common man. Blockchain thus helps to reduce the activities by bringing them to a common platform. Reduction in corruption: There are people those used to perform bad practices which involves brib- ing the government officers in order to buy or sell the properties. Blockchain based system will reduces the flow the black money from the system and everything will be legally done. Thus Blockchain based record system helps in reducing the corruption.

VIII. Future Scope:

Geographic information System (longitude & latitude): Geographic information sys- tem implies to storing land into the system with help of virtual reality using the coor- dinates longitude and latitude and Google API one can generate the real time land area virtually inside the system. This will help to recognize the land area when there are disputes related to land area between two land owners having lands side by side. Also it helps to recognize the land in case of any natural calamities.Multiple owner (family Tree): Multiple owner implies to land or a property having ancestral owner- ship and storing their family tree to the blockchain will help to ease of transfer of property in case of sudden demise of the owner who do not have pre-prepared will or power of attorney. Family tree will help to decide the next owner or the new owner of the property.

References

1. Victoria L. Lemieux : A typology of blockchain recordkeeping solutions and some reflec- tions on their implications for the future of archival preservation, 2017 IEEE International Conference on Big Data (Big Data).
2. Al Imtiaz , Md. Sazzadur Rahman: Integrated land information system: To ensure the transparency and availability of information, in context of Bangladesh, 2014 Internation- al Conference on Electrical Engineering and Information & Communication Technology, 14651107
3. Sourabh Paul ; Ujwal Kumar Durgam ; Umesh C. Pati: LANDSAT Enhanced Thematic Mapper plus image registration using SIFT,2015 Third International Conference on Image Information Processing (ICIIP)

4. "Why land tenure is important", FAO Corporate Document Repository, last ac- cessed on July 4, 2017,http://www.fao.org/docrep/005/y4307e/y4307e04.htm.

5. "India- Land Policies for growth and poverty reduction", Agriculture and Rural Development Sector Unit, India Country Management Unit, South Asia Region, World Bank,http://siteresources.worldbank.org/INTKNOWLEDGEFOR CHANGE/Reso urces/491519-1199818447826/landpoliciesforgrowthandpovertyreduction.pdf.
6. "Strengthening Arbitration and its Enforcement in India – Resolve in India", NITI Aayog, http://niti.gov.in/writereaddata/files/document_publication/Arbitra tion.pdf
7. Guaranteeing Title to Land, D.C. Wadhwa, World Bank,http://siteresources.worldbank.org/INTINDIA/Resources/dc _wadhwa_pape r.pdf.
8. Report of the Committee on Medium-term Path on Financial Inclusion, Reserve Bank of India, December 2015, https://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/FFIRA27F 4530706A41A0 BC394D01CB4892CC.PDF.
9. Operational Guidelines, Pradhan Mantri Fasal Bima Yojana, Department of Agri- culture, Cooperation and Farmers Welfare, Ministry of Agriculture & Farmers Wel- fare,http://agricoop.nic.in/Admin_Agricoop/Uploaded_File/OprationalGuidelines1822016.pdf.

Engineering knowledge as Graduate Attribute Attaining Assessment of Program Outcome

Dr. Rajesh Bansode, Dr. Sangeeta Vhatkar, Mrs. Pranjali Kasture, Mr.Vikas kaul
Information Technology, Thakur College of Engineering and Technology,
Mumbai, India
rajesh.bansode1977@gmail.com, vhatkarsangeeta@gmail.com, kasturepranjali@gmail.com, mvjvikas@gmail.com

Abstract – **Graduate attributes are the qualities, skills, and knowledge every graduate engineer needs to acquire and develop from time to time. Engineering knowledge is considered to apply knowledge of mathematics, science, engineering fundamentals and an engineering specialization to the solution of complex engineering problems. The accrediting bodies for engineering qualifications have developed outcomes-based criteria for evaluating programs through the graduates engineering knowledge. The engineering regulatory body that maintains quality of technical education across nation (AICTE) have developed the process of building competency-based standard with performance indicators for setting up question papers based on Blooms taxonomy which supports verification and validation of engineering knowledge gained by students which is tested through continuous evaluation. In the current paper the evaluation of Program Outcome /Graduate attribute-1 is done as Engineering Knowledge attainment which satisfies knowledge of computing, mathematics, science and engineering fundamentals for SE, TE & BE students with direct and indirect assessment tools.**

Keywords: Engineering Knowledge, Program Outcome, computing, mathematics, science and engineering fundamentals

I. INTRODUCTION

Engineering Program outcomes describe about the learning of students' measurement during their 4 years of graduation. The skills, knowledge students acquire in each semester is measured through two major modes namely direct & indirect tools respectively. The direct tools include continuous evaluation, end semester exam result, practicals' and assignments whereas the indirect tools include student validation, course and exit survey. Program Outcomes are related to outcome based education (OBE) which is student centric activity. Instructions during semester focus on measuring student performance i.e. outcomes. Outcomes include knowledge, skills and attitudes [1].

Outcomes focus on actions that signify student's learning by using concrete, measurable blooms taxonomy verbs. Benefits of OBE is the focus on outcomes that creates a clear expectation of what needs to be accomplished by the end of the course. Flexibility, Comparison, Involvement, Definition, Assessment problems, Generality, Involvement. Faculty members are responsible for learning and motivation among students to inculcate engineering knowledge through delivery of technical contents in practicals, theory lectures, tutorials, assignments etc [2]. Students use engineering knowledge to prepare for competitive exams through self-learning, collaborative & peer learning mode. The purpose of OBE is practiced by gathering and analyzing feedback of the knowledge and skills acquired by students.

The **advantage** of perculating engineering knowledge is that **outcome based education** teaches thinking skills and engages students in more diverse technical learning skills [3]. The learning skills include participation in national level technical events viz. coding competitions, hackathons, domain competitions etc.

The first program outcome (Graduate attribute) is based on engineering knowledge. It is about an ability to apply knowledge of computing, mathematics, science and engineering fundamentals appropriate to the discipline [4]. The courses under Information and Technology curriculum intend to deliver, the fundamentals of Set Theory Function and Relation, enable students to understand and apply the fundamentals to explain the concept of pigeon hole, recurrence relation and generating function, lattice, arithmetic modulo, Laplace transforms and Inverse Laplace transforms. The course also intends to deliver the fundamentals of data structures and analysis of various algorithms by providing a platform to learn, compare and apply different data structures in real world scenario. It is about fundamental knowledge of Digital logic, number system, conversions and Boolean algebra, apply and analyze the concept to create and troubleshoot a broad range of combinational circuits using digital ICs, flip-flops, counters, and registers, to prepare students to perform the analysis and design of various digital electronic circuits. It focuses on fundamentals of database management systems, understand & create Entity Relationship Model with application of relational algebra and basic SQL. Create systematic database and logical design through normalization,

recovery techniques and analyze query processing and optimizer. It emphasizes the fundamentals of object oriented programming features, building blocks along with constructors, and to learn the principles of inheritance, interface and exception handling mechanisms, multithreading and GUI Programming.

II. Related Work

The following section highlights different domain wise subject detailing with the objectives and expected outcomes. The engineering knowledge gained is by segregating the domains and detailing the required expected outcomes from each domain as mentioned in table 1. The engineering knowledge learning is diversified through various activities viz. conduct of technical seminars, industrial visits, activity based learning, bridge courses, online courses through NPTEL, MOOCs, Coursera, Udemy. etc.

Table 1-Domain wise objectives and expected outcomes

Domain Name	**Name of the Subjects**	**Domain Objectives**	**Expected Outcomes**
Software Programming & Development	Java Programming Lab, Operating System, Python Lab, Computer Graphics & Virtual Reality, Software Engineering &Project management, Software Design Lab, Cloud Service Design Lab, Cloud Computing & Services, Usability engineering , Ubiquitous Computing, Software Project Management, Software Architecture, Cloud Computing services, Computer Simulation & Modelling, Software Testing Quality assurance.	1. Build intelligent systems defining interactions with & within system. 2. Design self-managed (autonomous) systems based on awareness of context (environment). 3. Implement distributed system based on Smart DEI Model which in turn shall Quantify stakeholders & operations. 4. Construct systems which is pervasive while being distributed by hiding underlying operations and automating interactions. 5. Construct systems which shall be serving changing needs by being aware of resources & decide operation distribution in this multitenant system. 6. Develop Program applications for various types of operating systems using modern SDKs and Frameworks	1. Build systems which are intelligent while being capable of change flow of interactions. 2. Design context aware autonomous systems. 3. Implement systems based on Smart DEI model. 4. Construct system which appears to be one by defining transparent distributed environment. 5. Construct system which are scalable and incorporates multi features. 6. Design and implement applications for different devices including Servers, PCs, Mobile Phones etc.
Machine Learning & AI	Data structure & analysis, Applied Maths, Automata theory, Advanced data structures & analysis of algorithms, Image Processing, Multimedia systems, Intelligent System/Artificial intelligence	1. Devise systems that are intelligent. 2. Build systems following heuristic approach in operations. 3. Build system that learn from observation with and without human intervention. 4. Build system that can process one dimensional signals, image and video data.	1. Develop Systems which can take decisions, which can mimic human behavior and thinking 2. Build systems which can process and classify data on their own whether or not they know the outcome 3. Build system that can operate on electrical signals, image, audio and video data. 4. Build systems which are able to learn from the training data and improve their intelligence with every execution

Database Technologies	Database Management System SQL Lab, Advanced Data Management Technology, OLAP Lab, Data Mining & Business Intelligence, Business Intelligence Lab, Big Data Analytics, Geographical Information System	1. The students will be able to store, process, manage and retrieve the data from database. 2. The students will be able to segregate the data based on its worthiness. 3. The students will be able to understand the Volume, Variety and Velocity of the data. 4. The students will be able to understand the use of mapper and reducer in the HDFS. 5. Student will be able to explore the different technologies of database technologies.	1. The students will be able to fire the queries and manage the database efficiently. 2. The students will be able to differentiate between the relevant data and irrelevant data. 3. The students will be able to store the data in HDFS and process it using Map Reduce. 4. The students will be able to write programs on mapper and reducer in the HDFS. 5. The student will be able to identify various mini project, research based project ideas based on the database technologies.
Web Technology and E-commerce	Internet programming, E-Commerce& E-Business, Enterprise Resource Planning	1. Get knowledge of broad range of Internet tools 2. Design dynamic website using advanced web technologies. 3. Understand the concept of E-banking, electronic payment system 4. Describe different Business models, and develop an online Business Plan	1. Development of Web applications. 2. Introducing e-business applications in an organization; e-business technology and infrastructure. 3. Develop dynamic and interactive e-business applications and websites.
Information and Communication Technology	Logic Design, Principles of Communications, Digital Design Lab, Computer Organization & architecture, Computer Network, Networking Lab, Microprocessor Programming Lab, Cryptography & Network Security, Security Lab, Microcontroller & Embedded System, Internet Of Things , Robotics Wireless Networks, Sensor Network Lab, Wireless Technology, Storage Network Management & Retrieval, Wireless Sensor Networks	1. Develop an understanding of the data communication in networks, security issues related to data communication in networks and network applications. 2. Study comprehensive background in Network Systems and Data Communications procedures and latest technologies related to network security. 3. To Show students how to properly conduct research for recommending network and data communications hardware and software solutions. 4. To provide the skill sets to analyze, design, test, and evaluate network systems. 5. To Study concepts in Network Systems, network management and Data Communications Analysis and related fields. 6. To study and implement traditional as well as modern day computer networks - wireless and mobile, multimedia-based.	1. Describe security risks associated with computer networks and implement w3 mitigation strategies 2. Design and construct a network application that facilitates communication between computers on an intranet and the Internet. 3. Applies networking concepts to design, implement and maintain LANs and WANs to support modern implementations including internetworking and data convergence 4. Analyzes and applies security in computer and networking infrastructures while detecting any legal and ethical breaches 5. Demonstrates the management of IT infrastructures and projects

III. Methodology

Step 1- To test and measure the engineering knowledge based on academic results (2016-19 batch) as below**:**

Table.2 - Academic result (2016-19 batch)

May 17 (SE)	**End sem result**	**Dec 16 (SE)**	**End sem result**
Information Theory & Coding	97.39%	Analog and digital circuits	88.13%
Computer Networks	98.68%	Applied mathematics-III	82.29%
Automata Theory	99.34%	Data structure & algorithm analysis	98.04%
Computer Organization & Architecture	97.39%	Database management systems	95.45%
Web Programming	100.00%	Object oriented programming methodology	98.69%
Applied Mathematics IV	93.71%	Principles of analog and digital communication	92.86%
May 18 (TE)	**End sem result**	**Dec 17 (TE)**	**End sem result**
Advance Internet Technology	99.33%	Advanced database management systems	96.82%
Data Mining & Business Intelligence	100%	computer graphics & virtual reality	98.06%
Distributed Systems	99.35%	Microcontroller & embedded systems	95.54%
Software Engineering	99.35%	Open source technologies	99.36%
System & Web Security	100%	Operating systems	100.00%
May 19 (BE)	**End sem result**	**Dec 18 (BE)**	**End sem result**
Storage Network Management & Retrieval	100%	Software Project Management	99.35%
Big Data Analytics	100%	Cloud Computing	100%
Computer Simulation & Modeling	100%	Intelligent System	99.35%
Enterprise Resource Planning	100%	Wireless Technology	100%
Soft Computing	100%	E-Commerce & E-Business	100%
		Ubiquitous Computing	100%

Step 2- The process used for calculating the attainment **with H, M, L** (H-High, M-Medium, L-Low) is as follows:

1. H, M, L mapping for each subject has been taken depending on the relevance of CO's with PO's.
2. For each subject in all semesters the tools have been used with 80% weightage to direct tools like Semester end exam (SEE), Internal assessment test (IAT), Assignments, Practical's etc. and 20% weightage to indirect tools like course survey, student validation & exit survey.
3. The attainment for each subject is calculated and the final single value on scale of 3, as average for every subject is calculated.
4. The single value average of the subject is put up in the HML mapping of each subject as following:
   H: average value * 1
   M: average value * 2/3
   L: average value * 1/3

5. The average of all the subjects starting from the third semester to the eight semester is taken according to the H,M,L values in each of the subject's table for individual PO attainments.

6. After getting this attainment, 20% of exit survey is added to 80% of the above calculated average of semesters to get the final PO attainment

Step 3- Differential Evaluation: PO1 Attainment with result

PO 1 mapping for 2015-19 batch

Table 3- PO 1 attainment of S.E A.Y 2016-17 odd sem courses

Subject	ADC	DMS	DSA	Maths	OOPM	PADC
SEE	1.15	1.76	1.83	1.25	1.78	1.44
COURSE SURVEY	2.73	2.58	2.62	2.94	2.79	2.28
Avg	1.47	**1.93**	**1.99**	**1.59**	**1.98**	**1.60**
Total Avg.	**1.76**					

Table 4- PO 1 attainment of SE A.Y 2016-2017 even semester courses

Subject	AT	CN	WP	ITC	COA
SEE	2.19	2.40	2.25	2.34	2.16
COURSE SURVEY	2.58	2.82	3.00	3.00	2.94
Avg	**2.27**	**2.48**	**2.40**	**2.47**	**2.32**
Total Avg.	**2.39**				

Table 5- PO 1 attainment of TE A.Y 2017-18 Odd sem courses

Subject	MCES	ADMS	CGVR	OST	OS
SEE	2.28	1.55	2.13	2.12	2.12
COURSE SURVEY	2.81	2.77	2.77	2.78	2.76
Avg	1.93	**2.06**	**2.19**	**2.21**	**2.22**
Total Avg.	**2.12**				

Table 6- PO 1 attainment of TE A.Y 2017-18 Even sem courses

Subject	AIT	DMBI	DS	SE	SWS
SEE	2.60	2.42	2.74	2.80	2.69
COURSE SURVEY	**2.99**	**2.94**	**2.79**	**2.8**	**2.84**
Avg	**2.55**	**2.50**	**2.45**	**2.49**	**2.58**
Total Avg.	**2.51**				

IV. Result and Analysis

The final average P.O attainment is calculated for SE, TE and BE which is from above step 3

Program Attainment Class wise:

Table 7- Average PO1 attainment of 2015-19 batch

Class	PO1 Average Attainment		Average
	Odd sem	Even Sem	
SE	1.76	2.39	2.075
TE	2.12	2.51	2.315
BE	2.68	2.77	2.725

Table 8- PO 1 attainment with Direct and Indirect Tool (2015-19 Batch):

Direct tools	Final Attainment	Indirect tools	Final Attainment
Internal assessment test	**2.17**	Course Survey	2.47
End sem result	2.49		
Assignments	2.36	Exit Survey	2.76
Practicals	2.67		
Avg. Attainment	2.42	**Avg. Attainment**	2.61

The final attainment is carried out with 80% weightage of direct tools and 20% weightage from indirect tools. That is 80% of 2.42 and 20% of 2.61. Therefore the final attainment is 1.936 (direct tools avg. attainment) & 0.522 (indirect tool avg. assessment) totaling to **2.458** as final attainment**.**

Analysis:

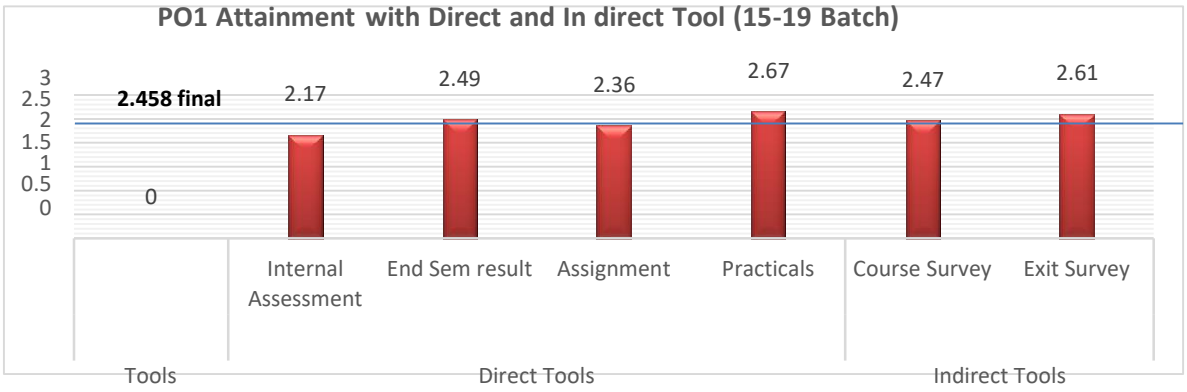


Fig.1 PO1 as Engineering Knowledge attainment using direct and indirect tools

V. Conclusion

The PO1 as engineering knowledge is proved to be beneficial concept of learning in attaining the PO attainment of 2.458 on the scale of 3 to mention that graduates can apply knowledge of mathematics, science, engineering fundamentals and an engineering specialization to the solution of complex engineering problems. Thereby, for better clarity the attainment is measured using direct and indirect tools for the students of batch 2015-19. The attainment tables indicate for students of batch 2015-19 is from academic year, A.Y 2016-17 from second year to final year of their core engineering department.

References:

1. M. Abdul Wahab Mohammad1 & Azami Zaharim1, Programme Outcomes Assessment Models in Engineering Faculties, Asian Social Science; Vol. 8, No. 16; 2012
2. Jennifer Hill, Helen Walkington & Derek France, Graduate attributes: implications for higher education practice and policy, Journal of Geography in Higher Education, 2016
3. Dr Patrick X.W. Zou, Associate Professor & Director of Research Students, Working Together to Achieve Graduate Attributes of Our Students, CEBE Transactions, Vol. 5, Issue 1, April 2008 pp 25-42 (18)
4. https://www.nbaind.org/files/NBA_UGEngg_Tier_II_Manual.pdf

Knowledge based tool for Academic Strategies and Analysis

Dr. Rajesh Bansode, Dr. Sheetal Rathi
Department of Information Technology, Department of Computer Engineering,
TCET, Mumbai

Abstract—**Educational institutions have a pivotal role in molding young, fertile and enthusiastic mindsets which in turn influences the society at large. This micro-chip society truly understands the importance and value of Knowledge which makes Knowledge Management (KM) essential for every educational organization. This can be identified and made use of at various levels such as administrative, academic (teaching and learning processes), research and student & staff welfare. Deployment of systematic KM enhances the quality of contribution made by an individual thereby encouraging them to co-operate and share their inputs, and to effectively challenge, negotiate and learn from others. In this paper we have applied knowledge management to categorically segregate students in low, medium and high categories using student validation tool. The results of systematic implementation of these categories on students evidently show that over the past four years, students go from low to medium to high category.**

I. INTRODUCTION

Knowledge Management is "A process by which the assimilated knowledge undergoes a cycle of knowledge generation, sharing and exploitation enabled by technology in order to achieve its objectives. The organizational goals are achieved through strategy-driven motivation and an approach providing knowledge to employees to enhance and use their capability to interpret data and information, and skills thereby infusing knowledge management practices into the daily work of the Academic Operation Area.

KM demands a mixture of skills, creativity, technical expertise and innovative capabilities of people; which form the base of KM. Knowledge is made available at three stages: before, during, or after KM-related activities.

Thakur College of Engineering and Technology has initiated KM activities at all levels viz. student, faculty as well as departmental level. In order to achieve results and outcomes to measure the innovative practices to convert them into best practices; all Departments including both academic (All branches) as well as Sections (Exam, T&P, Library etc.) have been assigned with different tools and tasks.

The case study discussed for KM is based on regular practice of student validation at department level. This validation is done at entry level of each year based on various parameters such as previous marks, attendance, term test, learning attitude etc.

The three fundamental pillars of knowledge management are as follows: (1) Knowledge Acquisition. The process of acquiring, gathering, developing and creating insights, skills, and relationships. (2) Knowledge Sharing. The process of disseminating and making accessible what is already known. A perfect example of knowledge sharing would be a decision support system that provides a novice teacher the best teaching practice by providing knowledge. Possibly, optimization of this sharing process must be done on the specific context; (3) Knowledge Utilization. Learning of broadly available knowledge is generalized, applied and integrated into the organization.

Effective knowledge management encourages the usage of more collective and systematic processes, which reduces our tendency to 'make the same mistakes' which otherwise is extremely costly and inefficient. Therefore, Effective knowledge management can enhance and dramatically improve quality of products and/or services. A complete, clear and thorough understanding of the stakeholder needs, customer needs, employee needs, industry needs, always has an obvious immediate effect on relationship management. Thus, accelerated knowledge creation, is the driver for innovation which forms the cornerstone of Effective Knowledge Management.

Steps taken to make proper use of acquired knowledge:

1) Faculty members are advised to come up with innovative practices every semester for overall development of students. The introduction of each of these innovative practices requires measurement forit is effectiveness which can be disseminated to all the departments if proven effective.

2) Suggestions from external stakeholders are very important for development which can be efficiently done by using various knowledge management tools.

3) Prediction of the possible problems that could arise if the internal/external benchmarking requirements are not met.

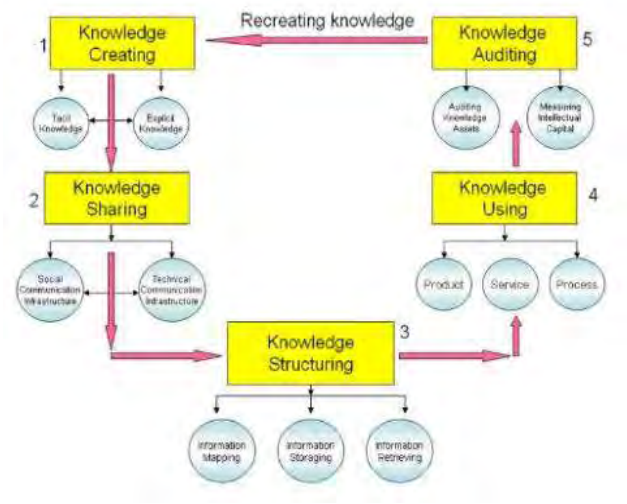


Fig 1. Knowledge Management Lifecycle

II. OBJECTIVE& SCOPE OF KM IN STUDENT VALIDATION

Knowledge is the key for decision making and strategy creation. Knowledge should transfer into an action but unfortunately it does not happen always. In order to sustain in competitive world all educational institutes should implement effective tools for knowledge management.

Following assignments should be activated for effective KM:

Mobilize the hidden implicit/tacit knowledge
Integrate knowledge from organization and make it accessible to all
Identify the missing knowledge
Create new knowledge
Make knowledge more accessible and usable
Create knowledge sharing culture to experiment and learn
Evaluate and reflect learning processes
Codify new knowledge.

KM helps educational institutes to improve their capacity of assembling and sharing of information and knowledge along with application of these to problem solving and support the research and repeated improvement of their work. KM of the educational system must reproduce and encompass information at all levels starting from management level to student level in order to improve professional knowledge of employees, to achieve quality of lecturers and students.

The Objectives of KM can be as follows:

1) To find the root cause of untraceable problems.
2) To foresee the requirements of Institute and be equipped enough for handling any dynamic requirements.
3) To develop optimal practices for holistic student development.
4) To inculcate all the suggestion and feedback from all stake holders into innovative and new practices.
5) To disseminate proper knowledge from time to time to concerned stakeholders.

III. DEVELOPMENT OF BEST PRACTICE

The practices of teaching learning which requires critical /problem solving skills, initially considered as innovative practice to categorize students learning based on their academic performance using continues evaluation with positive outcome as desired is considered as best practices.

The need to implement student validation as regular practice in teaching learning amongst students is to fully engage in their learning, for which validation of students is carried out to measure learning in 3 major categories High-H, Medium-M,LOW-L through which student diversity is measured. So is diversity among college students. The challenge is how to harness that strength, and how to unleash the creativity and exuberance for learning that is present in all students. Applying KM in student validation is one of the key factors for improving the categories of students who are lacking behind.

Student validation (SV) is considered for segregating students in 3 major categories based on their academic performance as High (H)- CGPA 7.75 and above, Medium (M)- CGPA ranging between 6.75 to 7.74 , whereas Low(L) – CGPA below 6.75. SV is prepared based on previous semester performance as SGPI, current sem.: term test marks, attendance, term work. The practice is carried out with a rigor practice to convert students from L to M & M to H category.

The Low (L) category students are informed to write remedial, assignment and practice the lab conduct to improve the academic performance.

The medium (M) category students are asked to update their technical skills by attending and gaining technical hands on training from the bridge courses and try to register and clear online certification courses viz. MOOCs, Coursera, NPTEL, Udacity etc.

The High (H) category student's performance is boosted by giving them inputs to clear the certification courses, internships, domain expertise, program specific research with placements etc.The measurement of student's growth signifies students performing better in the academics.

IV. IMPACT STUDY

The impact study provides an analysis of the student validation for the academic Year 2018-19.

Before: The student's performance was monitored categorically as H, M, L based on their academic performance. The institute followed the criteria of Mumbai University for A.Y 2018-19. The results are seen in fig. 2, fig 3, fig.4, fig 5.

After: In the A.Y 2019-20,TCET was awarded the status of Autonomy, due which there was improvementin student's academic which is reflected in their resultsas can be inferred. Due to various enlighteninginitiatives engaged by TCET such as Smart India Hackathon (SIH), technical seminars as per domain, professional bodies 'seminars& various workshops,students have imbibed the practical knowledge to exceltheir academics.

The measurement is done for consecutive batches where students' progression is perceived from the number of students shifting from L to M and M to H, whether there is a negative or positive trend in the growth of students. The respective processes are retuned as per students need, to meet higher benchmarks and try to get majority of the students in M, H category so that students are globally employable.

V. RESULTS

From the deployment of student validation, the faculty members get an idea of slow learners and can accordingly use various pedagogical teaching learning methods for such students. The figures below depict the growth in student's results from A.Y 2018-19 to 2019-20. There is a subsequent improvement in the results of students before and after the implementation of Autonomy for Information Technology & Computer Engineering departments.

Case A (IT Dept): Table -1: Student validation for A.Y 2018-19 Odd Sem-IT

Percentage of student in High , Medium and Low Category in A.Y 2019-20 odd sem –I.T			
Programme Avg	High (%)	Medium (%)	Low(%)
S.E	53	43	4
T.E	68	29	2
B.E	72	27	1

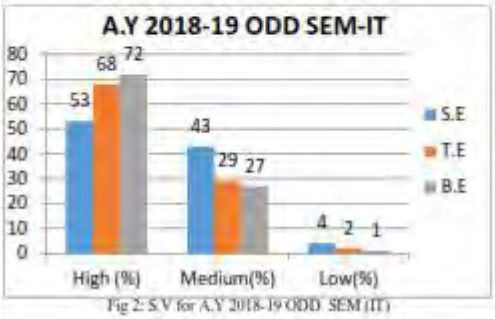


Fig 2: S.V for A.Y 2018-19 ODD SEM (IT)

Table -2: Student validation for A.Y 2018-19 EVEN SEM –IT

Percentage of student in High, Medium and Low Category in A.Y 2018-19 EVEN SEM -IT			
Programme Avg	High (%)	Medium(%)	Low(%)
S.E	54.83	40.91	4.20
T.E	67.01	30.20	2.40
B.E	71.92	26.70	1.38

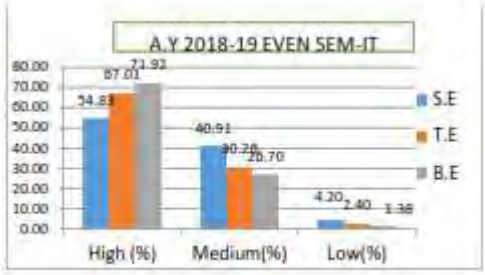


Fig 3: S.V for A.Y 2018-19 EVEN SEM (IT)

Case B(COMP Dept) Table -3: Student validation for A.Y 2018-19 ODD SEM-COMP

Percentage of Student Count in High, Medium and Low Category in A.Y 2018-19 ODD SEM -COMP			
Programme Avg	High (%)	Medium(%)	Low(%)
S.E	85.8	8.6	4.8
T.E	75.04	10.13	14.83
B.E	69.62	22.76	7.62

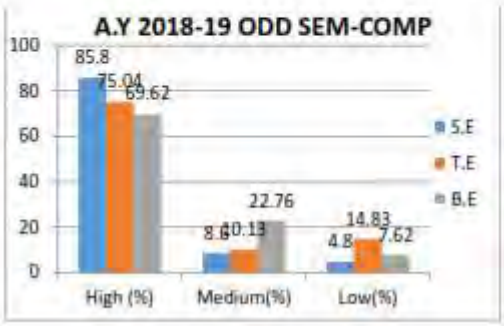


Fig 4: S.V for A.Y 2018-19 ODD SEM (COMP)

Table -2: Student validation for A.Y 2018-19 EVEN SEM –COMP

Percentage of Student Count in High, Medium and Low Category in A.Y 2018-19 EVEN SEM -COMP			
Programme	High (%)	Medium(%)	Low(%)
S.E	56.78	39.87	3.34

T.E	67.34	30.26	2.4
B.E	71.92	26.7	1.38

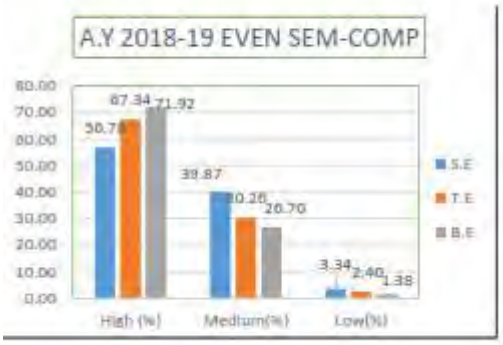


Fig 5: S.Y for A.Y 2018-19 EVEN SEM (COMP)

From the above demography, it is quite evident that the Student Validation Tool has proven to be the most efficient method of precisely gauging and continually measuring an individual's academic performance so as to measure/indicate overall growth of students. In order to convert students from L to M & M to H category for the overall improvement of student's performance at all levels (Curricular, co-curricular & extra-curricular), it refers to various minute calculations of each student categorize based on continuous evaluation (term work), internal assessment test, term work etc which gives a measure/ mechanism. The % of students moving from L to M & M to H are closely monitored for subsequent batches and inference is drawn for next subsequent batches about the various types of practices students to be given.

VI. CONCLUSION

In conclusion, the author has attempted to demonstrate the positive influence of Knowledge Management on studentteacher learning process and how effectively it has been implemented using a Student Validation tool to enhance oveall performance of an individual. The student validation tool is indeed, a foolproof system, however with asmall window for improvement for all the students irrespective of their category, asthe process to apply it on the students for measurement of their performance can be adaptive and not rigid,as it would vary based on various internal and external factors.

Evolution to Sustenance of Information Technology Academic Curriculum to Adhere to Industry Practices

Aniket Ladukar, Advait Maduskar, Divyam Choudhary, Dr. Rajesh Bansode
Department of IT, Thakur College of Engineering and Technology
ladukaraniket@gmail.com, advaitmaduskar.1@gmail.com, divyamchoudhary@hotmail.com, rajesh.bansode@thakureducation.org

Abstract – **The domain of Information Technology has been growing exponentially since the introduction of IT- Enabled Services in the industry since early 2000's. The IT infrastructure of any organization plays a crucial role in the processes of information management, processing, retrieval and communication. This article serves to provide insight into the radical growth of Information Technology over the course of the past few decades and to inspect the direction in which it is headed. Furthermore, the article also discusses about changes in Information Technology curriculum to bring sustenance as per industry practices. The current case study discusses about the paradigm shift in nature of IT-Enabled services which is reflected as per academic instructions being provided at the department.**

Keywords – Information Technology, Curriculum, Academia, Industry, IT-Enabled Services

I. INTRODUCTION

Information technology can be broadly defined as the use of computers, software (operating system/tools and application), communications, and networks to ensure that the information needs of an organization are being satisfied [1].

The department of Information Technology Engineering at the institute structures its curriculum in accordance with the industry trends pertaining to IT-Enabled Services and Applications. The undergraduate syllabus is compartmentalized into five sub-categories as domains and aims to provide a holistic technical education in a manner that promotes student interest in the domain. This part of the article is focused on analysing the domains into which the academic curriculum has been distributed.

The first of the domains is Software Programming and Development. The syllabus covered under this domain includes subjects such as Object-Oriented Programming, Structured Programming, Software Engineering and Project Management, among others. No doubt one of the most crucial skills required of an IT engineering undergraduate, Software Programming and Development serves as a basis on which to build skills that will allow students to specialize in any other software-related domains that they may be interested in.
The Department has 137 students registered under this domain with 9 faculties allotted to it. The subjects thoroughly cover the nuances required by students to gain a complete understanding of Software Engineering concepts as well as Process Models, both of which are fundamentals in any project taken on in the software industry.

The second domain is Artificial Intelligence and Machine Learning. Process automation and the usage of intelligent systems has seen an exponential rise over the course of past decade. Machine Learning and Artificial Intelligence are two of the technologies that are at the core of this boom. The demand for engineers who can build efficient intelligent systems – be it chatbots, intrusion detection systems, predictive models, or any such application – has seen an upward trend and is only likely to increase in the coming decade. Gartner Inc. has named 'Hyper automation' – a culmination of machine learning, packaged software, and automation – as one of the top 10 IT trends of 2020 [2].

168 students have registered for this domain and will be guided for 9 faculty members. The subjects covered under this domain serve to cover the basic concepts of machine learning and data mining. However, the depth of chapters in subjects such data mining and soft computing is often compromised to accommodate the vast breadth of topics. The converse of this is seen in subjects such as Data Structures and Algorithms. To address these shortcomings, the department has introduced electives such as Soft Computing and Advanced Data Structures and Algorithm Design. However, the lack of laboratory sessions for these electives leaves a need of practical understanding while being somewhat heavy on the theoretical aspects.

The third domain is Database technology. It forms the backbone of information technologies by defining storage and access methods. For a long time, the database industry has been dominated by giants such as Microsoft and Oracle with heavy focus on structured, SQL-driven databases. However, recent trends are indicating that the number of organizations that use hybrid databases (NoSQL + SQL) is rapidly increasingand the instances of pure SQL databases is steadily declining. The curriculum provides extensive training in the concepts of database management and SQL programming. Subjects such as Advanced DBMS also serve to introduce database security concepts along with the fundamentals of data warehousing. 194 students have opted for this domain and will be guided by 10 faculty members.

The fourth domain of Information and Communication Technology is an amalgamation of Telecommunications and Information technology. Its primary focus is to provide access to information using Internet, mobile phones, wireless networks and various other communication technologies. It encompasses all the devices and services which allow people and organizations to communicate digitally [3] [4].

We live in a world which is largely information driven. It has a great impact on our daily lives. Making all this information available to is the primary task of ICT. ICT has become an integral part of everyone's lives. For example, everyone with a cell phone and an internet connection can have access to enormous amounts of information right at their fingertips. Services such as Video conferencing, VoIP have turned out to be revolutionary in domains like Education, Healthcare etc. It has made it possible for people to obtain education from teachers present at remote locations. It has enabled patients living in remote areas to receive medical facilities such as consultation etc [5].

The curriculum offered under the domain of Information and Communication Technology covers a wide range of subjects that cover both wired and wireless communication systems. It would be apt to discuss that the syllabus, which covers computer networks, principles of communication, wireless networks, IoT among other subjects, is at pace with the requirements of the telecommunication industry. The 9 allotted faculty proficiently instruct the 145 students that have opted in for the domain while ensuring that no gap is left in the standards of the academic curriculum.

Overall, ICT is a platform that can be used for cross domain research that can involve the integration of database analytics, remote artificial intelligence, cloud- based data management, etc.

The fifth domain is Web Technology and E-Commerce. In recent years, with the increasing use of the internet, organizations are shifting from traditional native applications to web based services. This shift has been accompanied by the meteoric rise of web technologies for developing online interfaces for corporate services and research alike. Furthermore, the rise of E-commerce has boosted the need for web application developers in order to develop and maintain their portals.
The curriculum under Information Technology department covers subjects such as Internet Programming, Java Programming, ERP along with electives such as Advanced Internet Programming. A total of 243 students have chosen to opt in for this domain, making it the single largest category under the IT department.

While the syllabus covers topics such as basic HTML, CSS, JavaScript and PHP, it leaves out essential web development frameworks and libraries. As recent trends indicate, JavaScript based frameworks such as React, Node, Angular, among others are becoming increasingly popular tools in full stack development.

II. BACKGROUND

For the purpose of discussion, the growth and evolution of the IT industry can be viewed in three broad phases:

Pre-2000 era: The growth of software exporting firms.

For the two decades in this period, the software sector was largely comprised of firms looking to provide software services to global clients. The companies started solving Y2K issues for their customers and further extended their offerings to help companies manage their legacy portfolio of applications and infrastructure. The first wave of the global Internet and dot-com era created intercontinental Internet infrastructure. Indian companies were able to leverage this infrastructure to deliver software development- related services to global enterprises remotely.

Realizing the potential and the availability of talent, some multinational corporations established their own offshore development centers in India. Companies involved in the software aspects of hardware—for example, design of tools or VLSI (very large-scale integration)/system design—also took root, diversifying their services portfolio.

Circa 2000–2010: The rise of Indian software multinationals and R&D centers.

With experience in dealing with complex IT systems and confidence in working with international customers, several companies became multinationals with offices and centers across countries. They offered a wider range of services like executing large and complex projects involving integration, complete end-to-end solutions including management of IT infrastructure, running the services, providing IT strategy, and other related services.

Global multinational companies also realized India's potential in software services and started increasing their direct presence in India by setting up IT, business process management (BPM), and R&D centers. To date, 1,250 companies from around the world have set up their own centers in India across almost all key industry verticals. Software/Internet, telecom, semiconductor, automotive, and industrial are the top industries present, with R&D being a strong focal point. Enterprises across industries such as banking, retail, and healthcare also started driving digital engineering work from their India development centers.

Today, several centers have matured to deliver end-to- end products from India. These centers also act as the gateway to Asia, helping with product localization and creation of new products for these markets. Even next- generation companies have started setting up centers in India. Uber set up an engineering center in 2017, and OVH—a unicorn

from France that provides cloud services—set up an R&D center in the country last year.

Over 400,000 engineers work in global R&D centers in India. Bangalore, Pune, Hyderabad, National Capital Region (Delhi, Noida, Gurgaon), and Chennai are key locations for such centers, amplifying the possibility of ecosystem wide learning, relearning, innovation, and partnership.

Circa 2011 to present: Vibrant and innovation- driven multi-dimensional sector.

The Indian software ecosystem has now evolved into an extremely dynamic and varied sector that is building and managing the most complex IT systems for global enterprises. The combination of available talent, lower rates of brain drain to the U.S., the presence of large technology companies' R&D centers, and the presence of global venture capitalists has helped accelerate the growth of the start-up ecosystem. India, today, has over 7,000 start-ups (started less than five years ago), and over 1,200 technology start-ups were established in just the last year.

There are largely two types of technology start-ups. The first are consumer-led and largely focused on the India market. Initially these were replicas of U.S. companies, but soon morphed with unique innovations for the India market. For example, the cash on delivery model in e- commerce was pioneered in India and is now used globally. The second set of start-ups are focused on serving the U.S. and European markets.

In the last few years, 18 start-ups touched US$1 billion in market capitalization. Walmart bought India's largest e-commerce company, Flipkart, which is only about 11 years old, at a valuation of US$21 billion. OYO Rooms, a technology-enabled franchise model hotel chain, was started by a 20-year-old, and now has the largest number of rooms under management in India, overtaking both traditional Indian and global hotel chains.

Start-ups are driving innovation at an accelerated pace. To maintain the warp speed of innovation, large companies are building partnerships with the start-ups and are actively looking at acquisitions, both for talent and intellectual property.

For the purpose of discussion, the growth and evolution of the IT industry in India can be viewed as having occurred in different stages of development. According to an IBEF report on the evolution of the Indian IT sector, the process of evolution can be viewed in five stages, as shown in Fig. 1 [6].

According to a NASSCOM report titled “Industry Performance: 2018-19 and what lies ahead”, the past year was defined by technological trends such as Digital Acceleration, which led to a 33 billion USD increase in digital revenue with a 30% year-on-year growth. The same report stated that over 7200 new tech start-ups were launched, which resulted in 8 unicorns over the past year. Furthermore, a 6.6 billion USD investment in E-commerce firms that cater to the demands of 120 million customers lead to an additional boost of 43 billion USD to the digital revenue channels. The government provided extensive support to sustain this technological growth by announcing a nationwide AI for

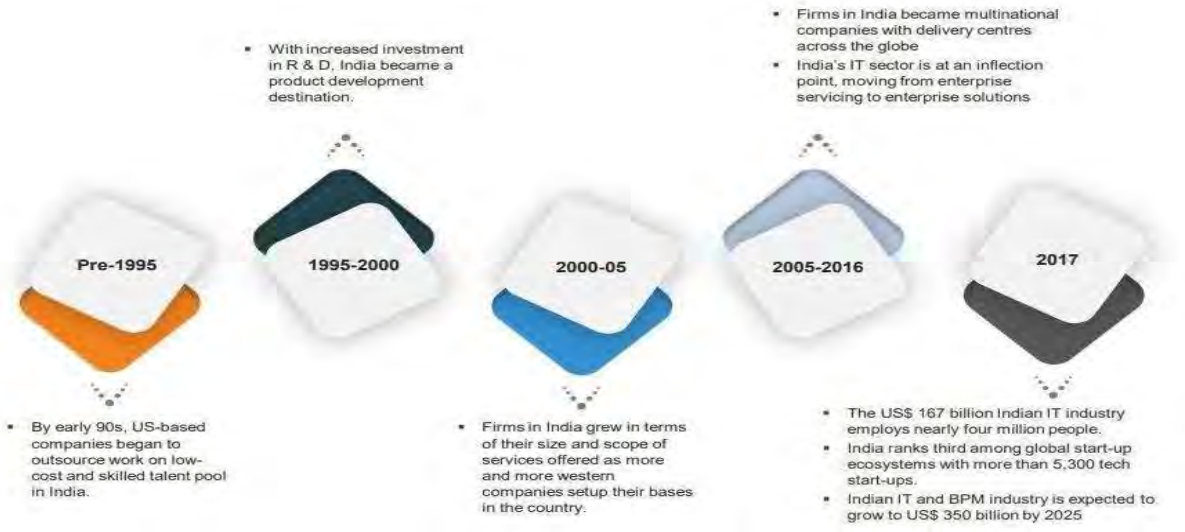


Fig. 1 Evolution of IT industry

all initiative. As a direct result of this move, 400 automation and artificial intelligence start-ups saw light in the past year [7] [8].

NASSCOM also defined evolving digital metrics for the measurement of the growth of evolving technologies and for the recognition of upcoming trends [9]. They are:

Customer Satisfaction: Measuring customer satisfaction emerges as a key metric for Digital growth
o Metrics: CSAT Scores
Investments in Digital: Investing into building capabilities on Digital platforms and services.
o Metrics: Investment in CoEs, Domain
Talent, IP
IP Monetization / Patents Filed: Increased focus on creating products / platforms on new age technologies
o Metrics: No. of Products / Platforms created, Patents filed
Digital Revenues / Margins: Measure and disclose Digital revenues
o Metrics: Growth in Digital revenue
Client Mining: Actively mining targeted clients to drive Digital revenue (E.g. Targeting clients based on spending propensity & industry).
o Metrics: Large deal tracking, Digital Share of the deal
Driving Automation Internally: Deploying automation to boost productivity.
o Metric: Bots deployed
Digital Skilling: Re-skilling existing employees and disclosing percentage of digitally skilled employees.
o Metric: Digital Workforce

According to official sources, the fastest growing trends in the technological industry in India are Artificial Intelligence and Internet of Things. This is supported by the data for the number of patents filed in India, between 2015 to 2018. The greatest number of patents were filed under the domain of AI, with a total of 330 patents filed under it. AI was followed by Cybersecurity, with a total of 193 patents registered. The third domain under which most patents were filed was IOT, with 107, and finally, 88 patents were filed in India for Cloud Computing.

Furthermore, the Indian Brand Equity Foundation released a report [6] that highlighted notable trends in the IT industry, identifying them as follows –

Global delivery model: Indian software product industry is expected to reach the mark of US$ 100 billion by 2025. Indian companies have set up over 1,000 global delivery centres in about 80 countries over the world

Leading sourcing destination: India is a prominent sourcing destination across the world, accounting for approximately 55 per cent market share in the global services sourcing business, as of FY18. India acquired a share of around 38 per cent in the overall Business Process Management (BPM) sourcing market.

Rise of Onshoring: Fall in automation costs and rise of digital has led to higher onshoring by the industry. Onshore revenue of Indian IT industry has grown from around 48 per cent in 2011-12 to 55.2 per cent for the quarter ended June 2018.

New technologies: Disruptive technologies, such as cloud computing, social media and data analytics, are offering new avenues of growth across verticals for IT companies. The SMAC (social, mobility, analytics, cloud) market is expected to grow to US$ 225 billion by 2020

Rural Development: The National Optical Fibre Network (NOFN) is being laid down in phases to connect all the 250,000 gram panchayats in the country.

Changing business dynamics: India's IT market is experiencing a significant shift from a few large- size deals to multiple small-size ones. The number of start-ups in technology is expected to reach 50000, adding to around 2 per cent of GDP

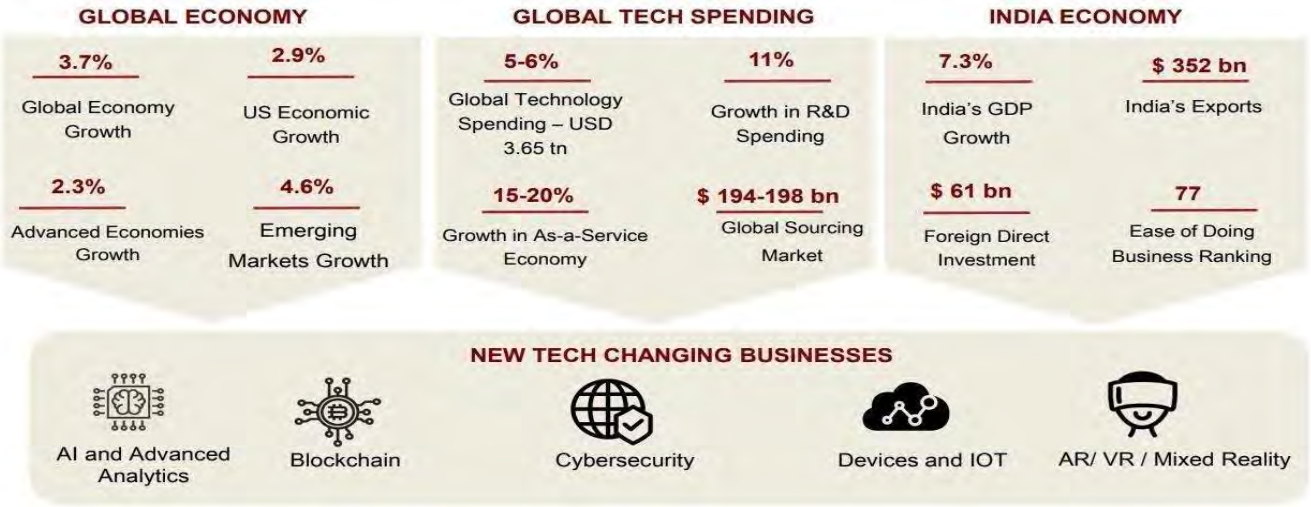


Fig. 2 Review of the last Financial Year in the Indian IT Industry

III. DOMAIN TRENDS

A. Software Programming and Development

The decade from 1990-2000 saw the rise of networking-based applications - the Internet. This led to a very rapid surge in demand. The ascent of the Internet led to a very rapid surge in the need for international information display/e-mail systems on the World Wide Web. Developers were required to deal with representations, maps, photos, other image formats, some animations like never before, with only a few known methods to optimize image display/storage (using thumbnail image). With the rising demand for software in many smaller organizations, a need for low-cost and economical software solutions was created in the early 2000s.

The demand was met by the growth of simpler, accelerated methodologies that developed running software, from requirements to deployment, agile & more accessible. The decade saw evolution of rapid- prototyping to an entire lightweight methodology, such as Extreme Programming (XP), which attempted to facilitate many areas of software engineering, including requirements gathering and reliability testing for the growing, vast number of small software systems. Software engineering is a budding and developing discipline. As per trends, the Software Engineering and Project Management evolved along these years will be relevant for the next decade as well. We see the trend of different programming languages and frameworks popping with each having their own speciality.

The present IT curriculum helps the student by providing training in Java, Python, Web Programming, Software Engineering and Project Management, Software Process Automation, Software Testing and Quality, UI/UX Design and Game Programming. Here DevOps, Functional Programming are few of the new disciplines that have emerged and hence been accordingly introduced in the curriculum [10].

B. Artificial Intelligence and Machine Learning

Significant advances in all areas of AI were made during 1990s, with important demonstrations in intelligent tutoring, machine learning, multi-agent planning, case-based reasoning, scheduling, data mining, uncertain reasoning, natural language understanding and translation, virtual reality, vision, and games. The early 21st century saw a boom in intelligent systems and robotics. The century witnessed the introduction of Interactive robot pets, humanoid robots. Humanoid robot capabilities were made to be able to walk as fast as a human. Search Engines saw breakthrough results with the use of Recommendation technology and Natural Language Processing. AI was introduced in understanding biological intelligence, even driving autonomous cars.

The decade from 2010 introduced virtual assistants - Siri, Google, Cortana, Alexa. Chatbots became an integral part of lives and data became the new oil. From
2010, the AI/ML/DL domain began to see applications in all aspects of life - healthcare, agriculture, transport & logistics, manufacturing, Auto & Aero and Utilities. The domain of AI has given rise to a new set of talents that are able to work with Data science toolkits (R, NumPy, Pandas, Scala-ML), Open source (Docker, Kubernetes, Helm, Linux, Horizon, etc.), AI/ML techniques and algorithms. These are technologies which promise to stay relevant for the next decade. Accordingly, college has introduced the relevant disciplines in the form of compulsory or elective subjects. They include Applied Mathematics, Python and Advanced Python, Computer Graphics and Virtual Reality, Data Mining and Business Intelligence, Big Data Analytics [11].

Fig. 3 Trends in Software Programming and Development

Fig. 4 Trends in AI/ML

Fig. 3, Fig. 4, Fig. 5, Fig. 6 and Fig. 7 signify
* Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means that there was not enough data for this term. [18]

C. Database Technology

The rise of the Internet applications led to the exponential growth of the database industry. Naive desktop users began to use client-server database systems to access computer systems that contained legacy data. The use of CGI, GCC, MySQL, Apache, and other systems brought open source

solutions to the Internet. This marked the dawn of Online Transaction Processing and Online Analytical Processing Technologies. The Data Warehousing Technology rose in demand. Despite the Internet industry experiencing a decline in the early 2000s, database applications continue to grow. Microsoft, IBM and Oracle tapped in this market trend becoming the current leading database companies.

Databases are now used to enhance our daily life. Cloud storage services, weather forecasting are applications of advanced data management technologies. The current relational databases include giants such as Oracle, MySQL, and DB2. We see a trend that focuses on making newer, potent technologies available to everyone. Database technologies alternative to RDBMS like Hadoop and NoSQL have developed. Database domain goes hand-in-hand and serves as an important prerequisite for Web as well as AI/ML domains. Data scientists expect the leading cloud database platforms to address the broader range of workloads and provide capabilities that these alternative database models have enabled.

Thus, Database modelling seems to be relevant in further coming years being a very specialized field of study. The students are taught Data Structures and Algorithms, Database Management System, SQL/PLSQL, Distributed System, Cloud Computing & Services, NoSQL, Storage Area Networks, Management Information System. While most of the concepts remain constant due to the legacy nature of the discipline, we see the inclusion of new NoSQL Technology in the IT curriculum which is seen as the rising trend in the database industry for the decade [12].

D. Information and Communication Technology

The world wide web became accessible to everyone in 1991. This led to a huge transformation in how people interact with each other and their surroundings. The internet facilitated the creation of information superhighway. The domain of ICT was powered by disciplines of Database, Web Technology and Software Engineering. Cell phones became as common as the computers in the early 2000s and the number of cell phone users grew from 12.5 million to 1.5 billion while the number of internet users grew from 2.8 million to 631 million in just a decade.

Over 2 billion people used the internet by 2012 which is twice the number using it in 2007. Cloud computing became mainstream by 2010. Half of the world's population was connected by 2016. The decade from 2010 witnessed an exponential growth IOT based devices. Daily objects started to include some form of micro computing. This also meant that these devices had all the security faults of computer systems as well.

Thus, the discipline of Cryptography and Network Security gained a small momentum which further developed into high demand jobs. Smartphones became more feature rich and more accessible. The Apple iPhone and Google's Android operating system pushed the smartphone into the mass market. Evolution of wireless technology from 2G to 5G brought faster data transfer speeds than other networks. Voice over Internet protocol (VoIP), sites like Facebook and Twitter allow users to communicate with networks of people.

The field of ICT currently has IOT as its most promising domain. The discipline of IOT being empowered by AI/ML and database has created a new discipline of Edge Technology. This new technology has created a variety of technical expertise in demand like Wireless, LTE, Bluetooth & Wi-Fi, 2G/3G/4G/5G –connectivity, protocols, security, Network infrastructure, Management, Architecture, Data Protocol (TCP/IP, SSL/TLS, DNS). The IT curriculum has adapted the required technology by providing courses on Digital Circuit Design, Principles of Communication, Computer Architecture, Computer Networks, Cryptography and Network Security, Distributed Systems, Cloud Computing Services, Microprocessor, Microcontrollers and Embedded Systems. The Curriculum has also newly included Internet of Things as a discipline for students to be able to adhere with changing industry practices [3] [5] [13].

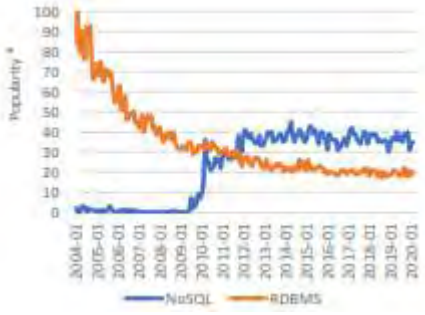


Fig. 5 Trends in Database Technology

magenta and other MVC Frameworks which are industry standards. All these subjects can be included in the curriculum to align it with the industry practices [14] [15] [16].

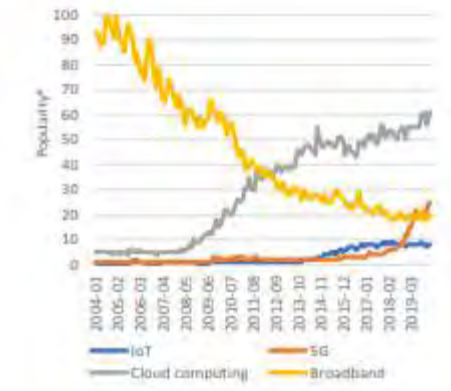


Fig. 6 Trends in ICT

E. Web Technology and E- Commerce

The decade of the 90s witnessed the rise of the Internet. The first web browser-Worldwide Web and search engine-Archie was created. HTML, flash, JavaScript became the go to scripting languages and are still at the top. This decade was a decade of experimentation and finding new applications of networking through the internet. In the later decade, W3C released new standards for HTML (HTML5) and CSS (CSS3). Development in Database technologies enabled new API Technology which help develop application interfaces for mobile technology.

The rise of handheld devices and smartphones gave rise to a new industry of Mobile App Development, Mobile- Friendly UI Design. Advancements were made in UI design with the introduction of MotionUI. Boom of ecommerce websites mandated online presence of a business since it started bringing the major business opportunity. Now, web technology includes AMP which is designed to enhance the overall performance of web content and ads while improving website speed and user experience, Progressive Web Apps , RAIL Concept, Motion UI, Browser Extensions , VR And AR, Single Page Applications, Internet Of Things, Blockchain Technology, better libraries and frameworks, Search Engine Optimization, Rich Internet Applications.

The IT Curriculum meets the industry requirements by including disciplines of Web Programming, ADO.NET, Advanced Internet Programming, Web Commerce, UI/UX Design. The curriculum has now adapted to newer technologies and methodologies of Blockchain as well Ethical Hacking and Cyber Security. On the other hand, the curriculum also lacks training on advanced frameworks like Laravel, Django,

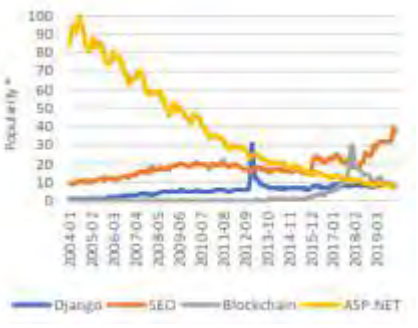


Fig. 7 Trends in Web Technology

IV. CURRICULUM DESIGN

The Second Year (S.E.) curriculum includes Program Specific core subjects such as Data Structures and Algorithms, Computer Organization and Architecture, Digital Circuit Design, Database Management Systems, Java Programming, Python Programming, Advanced Python Programming, Computer Networks and SQL/PLSQL along with laboratory sessions and project based learning sessions which would introduce to them the core areas of the particular course giving them in- depth knowledge and form the basic foundation in them.

The fifth semester offers Domain Specific "Industry Electives" such as Software Architecture, Soft Computing, Distributed Systems and Basics of Machine Learning, along with core subjects including Operating Systems, Cryptography and Network Security, Automata Theory and Web Programming, which satisfy the current Industry demands and requirements. The student's knowledge enhances and makes them abreast with the current technology.

The syllabus scheme provides credits for online courses from Semester VI onwards. This is to motivate the students to enhance their knowledge and encourage Self-Learning amongst students. Furthermore, the sixth semester has in its curriculum subjects such as ML and AI, Software Engineering and Project Management along with electives such as DevOps, Advanced Data Structures, Cloud Computing and Services, Data Mining and Business Intelligence, IOT, Advanced Internet Programming, NoSQL, Image Processing among other, in order to better familiarize students with industry-level knowledge of the technologies under implementation and to train them towards becoming more versatile.The Final Year, Bachelor of Engineering (B.E.) curriculum allows students to choose their own subjects by taking up professional electives and open electives from a comprehensive list of options that include Blockchain, Ethical Hacking, Big Data Analytics, Web Commerce, Cybersecurity and Law, Data Science, UI/UX Design, Game Development, MIS and Research Methodology. This scheme is aligned towards Generalized Knowledge which is an important component of crystallized intelligence. It would help students grow both on personal as well as academic level and develop the openness to experience and analyse situations for better solutions.

Throughout the curriculum, the students are not limited to studying the core subjects prescribed by the University. They are offered a number of Professional as well as Open electives which they can opt from as per their inclination. The students are also awarded credits for complete online courses in various domains with an aim to inculcate a habit of self-learning among the students.

Each semester is divided in two teaching schemes, namely – Program Specific and Holistic Student Development. The Program specific teaching scheme focuses on covering the core as well as domain specific electives in order to ensure that the students have a strong academic foundation. The Holistic Student Development scheme offers Professional skills, Project based learning and activity-based learning. Professional skills ensure that students are taught tools and

frameworks currently used in the industry in order to keep them up to date with industry requirements. The Project based learning provides them an environment to work in a team and learn various aspects of working collaboratively and enhance their team building skills. The Activity based learning provides them an avenue to get practical hands on experience of the skills learned throughout the coursework.

Overall, the curriculum has been designed after a thorough inspection of the trends in the IT sector over the past three decades and by analysing them to identify the direction in which the IT industry is headed over the coming years. This revised curriculum takes into consideration the changing needs of the IT industry and is so designed to provide students with a strong platform to hone their fundamentals and inculcate them with critical skills that will help them succeed as professionals. At the same time, by promoting research- based and project-based learning, the students are

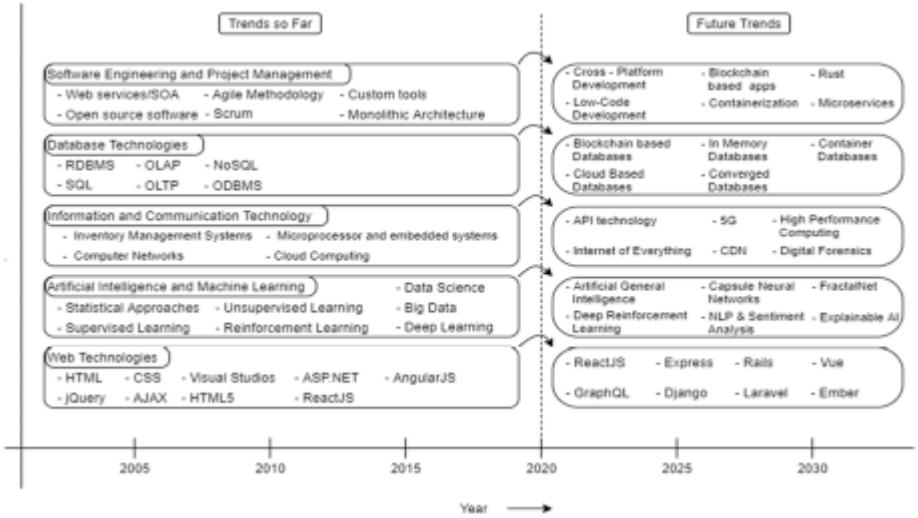


Fig. 8 Industry Trends

Fig. 9 Curricular Aspects of Academic Program

enabled to adapt to any changes by reskilling themselves as per the need of the industry.

V. FUTURE SCOPE

The theoretical instructions and concepts are introduced to students in proper depth. However, there is still lacking in the practical training provided.

One of the observed shortcomings in Programming laboratory sessions is that the practical sessions are geared towards strengthening the students' syntactical understanding of the language and not on actual practical implementations such as problem solving.

While the software engineering curriculum is up to date, modern concepts such as Git and tools such as GitHub which are widely used in industry should be introduced to students as a part of the practical sessions.

The lack of lab sessions for elective subjects such as Advanced Data Structures and Algorithms leaves the need for practical understanding of concepts. Introducing more rigorous lab sessions can help familiarize students with writing code as well as solving the type of problems that are presented to them in interviews.

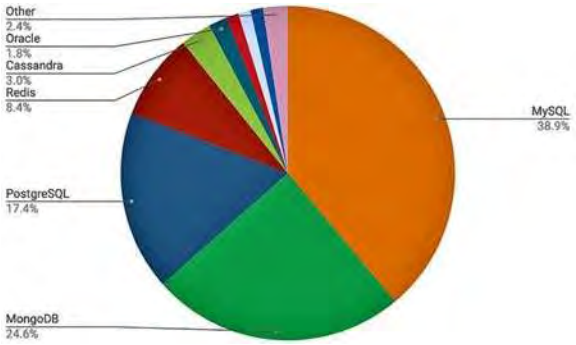


Fig. 10 Popularity of databases in use

As seen from the Fig. 10 above [17], it is clear that there is a demand in the IT industry for database engineers with knowledge of NoSQL databases such as MongoDB. The curriculum would do well to provide a strong introduction to NoSQL and Hybrid Databases, along with a more thorough explanation of data warehousing concepts.

As databases are migrating to the cloud, it is also of essence to include in the syllabus cloud computing concepts at a practical level. While the subject remains a part of the syllabus, it only does so in a theoretical capacity. Including practical applications such as cloud application development, docker, Kubernetes, etc. would enable students to get a hands-on understanding of classroom concepts.

Splitting the Internet Programming curriculum over two semesters in order to accommodate frameworks such as React, Node, Angular, etc., which are widely used in industry, as a part of the syllabus would help the department train students better. The curriculum should also include training in frameworks such as Django, Spring MVC, Ruby on Rails, etc. Given the depth and the breadth of the domain, the most viable path going forward is to allow students to choose a domain to specialize in within the department and provide extensive training in the domain to those that have opted in.

The institute has adopted a credit-based grading system. A more flexible course structure that would allow students to pursue more courses in the domains that interest them, while earning credits for the same should be implemented in order to promote specializations. The practical examinations, too, which rely heavily on oral examinations, should be replaced with project-based scoring, which will ensure student engagement throughout the semester while introducing them to new concepts in the respective subjects. While the department has largely adopted this system, the implementation relies on practical write-ups and vivas as rubrics for student scores. The credit-based system should be rigorously adopted at a practical level, wherein credits are awarded by their performance in the lab, with unique, original work as the output.

References

[1] C. Tupper , Data architechture: From Zen to Reality, Elsevier, 2011.

[2] K. Costello and M. Rimol, "Gartner Identifies the Top 10 Strategic Technology Trends for 2020," 21 October 2019. [Online]. Available: https://www.gartner.com/en/newsroom/press- releases/2019-10-21-gartner-identifies-the-top- 10-strategic-technology-trends-for-2020.

[3] P. Christensson, "ICT Definition," 04 January 2010. [Online]. Available: https://techterms.com/definition/ict.

[4] M. Rouse, "ICT (information and communications technology, or technologies)," July 2019. [Online]. Available: https://searchcio.techtarget.com/definition/ICT- information-and-communications-technology-or- technologies.

[5] "Importance of Information and Communications Technology (ICT) in our daily life," 29 November 2016. [Online]. Available: https://pcdreams.com.sg/importance-of- information-and-communications-technology- ict-in-our-daily-life/.

[6] IBEF, "IT & ITES," [Online]. Available: https://www.ibef.org/download/it-ites-feb- 2019.pdf.

[7] NASSCOM, "Industry Performance: 2018-19 and what lies ahead," [Online]. Available: https://www.nasscom.in/sites/default/files/Indust ry-Performance2018-19-and-what-lies- ahead_0.pdf.

[8] NASSCOM, "Future Skills : An Industry Response," [Online]. Available: https://www.nasscom.in/sites/default/files/Future Skills_An_Industry_Response.pdf.

[9] NASSCOM, "Emerging Technologies," [Online]. Available: https://www.nasscom.in/system/files/secure- pdf/emerging-technologies.pdf.

[10] "History of software engineering," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/History_of_softwar e_engineering.

[11] "Timeline of artificial intelligence," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Timeline_of_artific ial_intelligence.

[12] "A Timeline of Database History," Quickbase, [Online]. Available: https://www.quickbase.com/articles/timeline-of- database-history.

[13] J. Strickland, "What is the future of communication?," HowStuffWorks, 12 April 2010. [Online]. Available: https://electronics.howstuffworks.com/everyday- tech/future-of-communication1.htm.

[14] "20 years ago today, the World Wide Web opened to the public," TheNextWeb, 6 August 2011. [Online]. Available: https://thenextweb.com/insider/2011/08/06/20- years-ago-today-the-world-wide-web-opened-to- the-public/.

[15] N. Balog, "Top 8 Web Development Trends of 2019," CodingDoJo, 23 January 2019. [Online]. Available: https://www.codingdojo.com/blog/web- development-trends-of-2019.

[16] K. Faber, "The Future of Database Technology," Innovation Management, 9 May 2018. [Online]. Available: https://innovationmanagement.se/2018/05/09/the -future-of-database-technology/.

[17] K. Anderson, "2019 Database Trends – SQL Vs. NoSQL, Top Databases, Single Vs. Multiple Database Use," 06 March 2019. [Online]. Available: http://highscalability.com/blog/2019/3/6/2019- database-trends-sql-vs-nosql-top-databases- single-vs-mu.html.

[18] "Google Trends," Google, [Online]. Available: trends.google.com.

Judicial System Using Blockchain

Vinay Agarwal, Prartish Bhansali, Mahir Bhatt
Thakur College of Engineering and Technology Mumbai, India
agarwalvinay180@gmail.com, bhansalipratish287@gmail.com, mahirbhatt9@gmail.com

Abstract— **Blockchain is an incredible invention, the brainchild of an individual known by the pen name, Satoshi Nakamoto. But since then, it has evolved into a technology which holds a vision greater and better than the greatest of creations in the field of technology. A Blockchain is an incorruptible and non- tamper able public digital ledger of data collected through a network. It is how this information is recorded that provides Blockchain its ground-breaking potential. Thus, Blockchain technology provides integrity, authentication, confidentiality and security to everyone on the chain.**

The Blockchain network has no central authority or governance which is the very definition of a decentralized system. Since it is a distributed and an immutable ledger, the information in it is open for everyone to view. The legal industry has ample of features togain from the Blockchain by integrating their existing software with Blockchain.

One of the main issues of judicial systems today is the management and handling of the evidences. From the time of collection by the court it undergoes a series of ownership and scrutinizing and may lead to manipulations that may be deceitful. Also, the current process is entirely paper based processes and acting on single database system which makes it a single- server model. All the data being stored on to a single server makes the system vulnerable on being prone to attacks and can lead to permanent data loss and not to mention making the process slow. For that purpose, this project deals with successfully embedding Blockchain into the judicial systems for making it even faster, secure and making the people possessing the evidence accountable for their deeds which ultimately leads to provide equal justice to all. Our prototype is entirely developed on the Ethereum private network

I. INTRODUCTION

The world is experiencing a great pace of globalization and digitalization. An exponentially increasing number of transactions are being carried out online between entities across the world. If the awareness and implementation of the Blockchain technology becomes a world-wide reality, in a not so distant future, most goods, labour and capital will be allocated through decentralized global platforms and thus make the large amount of systems non-tamper able and error free.

The Blockchain is a digital ledger that logs transactions like a book-keeper that can be programmed to record not just financial transactions but virtually everything with a value. This ledger system is of great use to our proposed system, since all the transactions will be logged in the chain, we as a user can make a proper history of the task, which is the most crucial element of our system.

The legal system is currently paper-based that may result in loss of evidences, duplication, errors, which makes the entire process time consuming and may lead to injustice on the path of an innocent. Blockchain could shape law as much as law shapes them. Smart contracts are automatically executed as they are programmed, but it cannot analyse subjective judgments or include elements from outside the Blockchain. Existing dispute resolution technologies are too slow, too unreliable and too expensive for a decentralized global economy operating in real time. A reliable, transpare-
-nt and decentralized and accountable dispute resolutionmechanism that renders ultimate judgments about the enforceability of smart contracts, key mantra for the Blockchain era.

The sole purpose of our system is to transform the way in which the judicial entities store the data and the evidences, in order to make the system transparent and accountable and tamper-proof.

Accountability: None is held responsible for delay in current Indian Judicial System. Even though court officials might be justified in being overburdened, but finally it's the common man getting pissed off.

Provenance: Provenance of data is lacking currently in Indian Judicial System. Originals can be easily forged and hence, questionable.

A. Summary:

II. HISTORY

Transparency: Despite having privileges of certified copy for legal documents, Indian Judicial System lacks transparency mostly due to malpractices conducted by court officials.Judicial System is an important part of democracy and plays

a crucial role in how democracy works. In current Indian judicial system, there are two ways that a common man can approach court viz. via advocates or through court officials as represented in figure 1.

Fig. 1. *General Process to Approach Judicial System.*

In both the cases, legal procedure is an expensive affair. Common public has no or very limited knowledge about legal proceedings. They completely rely on advocates or court officials for any legal proceedings. There might be various reasons for same such as no legal knowledge, illiteracy, too busy or no interest to gain knowledge regarding legal process and so on. This blind trust on intermediaries might result in expensive dealings or in extreme case, cheating.

Current Challenges:

Below listed are the prominent challenges faced by current
Indian Judicial System:
Data Integration: Currently judicial officials have data records of only respective zones. Integration of criminal records across country will have better opportunity to track criminals and their past records to restrict unfortunate events in future.
Scalability: Current judicial system is not scalable enough to maintain record of cases across state.

B. Background:

In the recent years, with the emergence of the concept of Blockchain and its early use of secure and peer to peer transfer of money in terms of crypto currencies has been one of the widely witnessed phenomena globally. Ethereum was setup based on the Blockchain technology which made it efficient and secure for developing decentralized applications by setting up a private network on determining the consensus. There are about 3 crore cases currently pending in Indian courts, including 42.5 lakh cases in high courts and 2.6 crore cases in lower courts. Even if 100 cases are disposed of every hour, it would take more than 35 years to catch up. For this purpose, this project deals with creating a smart contract that will overcome the drawbacks of the current judicial system using Ethereum networks for fast judgments, reducing the paper work and bringing transparency in the overall system.*C. Problem Definition:*

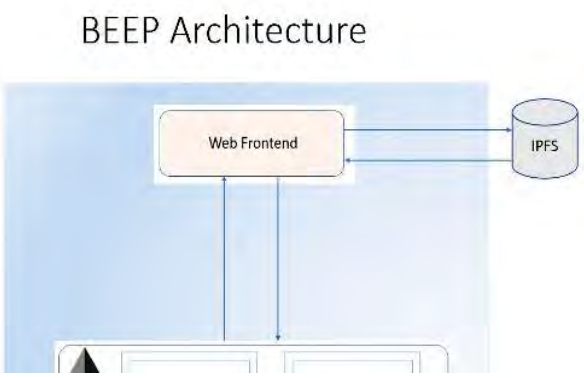
In the current law and order, the system is purely paper based processes. This might result in the loss of evidence, duplication, tampering, and errors and may lead to injustice on the path of an innocent. Using traditional database system can definitely pose a problem in future as it can be hacked, also it would not be possible to access the data when the server is down. Single point of failure can lead to permanent data loss which can be fatal for data critical tasks. The existing system has no facility to provide versioning of the evidences and is not immune to evidence tampering.

D. Proposed Aspects of the Project:

The proposed system will consist of a private Ethereum network connected to different nodes. The data will be stored on a decentralized file storage system (Interplanetary File System) and the hash of the evidence will be stored on the chain. It consists of python flask as a micro-framework. It will allow the user to create and upload the evidence and also transfer its ownership to another user in the network. The user will be the able to see the timeline of the evidence's custody in the update chain of custody module. These evidences will be stored in the IPFS and the hash associated with the evidences will be stored in the private Ethereum Blockchain. The evidences can be retrieved using the hash from get evidence module.

Ethereum private network is to be set up as fully public and uncontrolled network might cause delay in committing transactions. By setting up private network it makes it simple and fast for the users to create evidence and upload them with minimal gas consumption and also causes no delay. The smart contracts are written in solidity language. The interface is designed using Flask which is python- based framework. Python flask used for designing the frontend can be used for storing files and retrieving hash of the files uploaded in IPFS and that hash is uploaded in the network by using web3.py

III. SYSTEM ANALYSIS

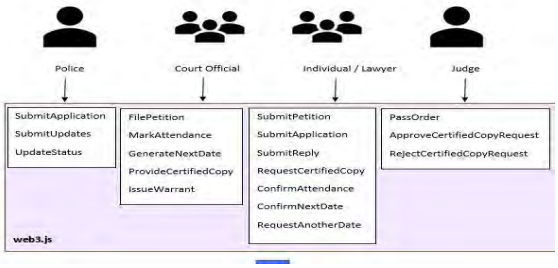


The user once registered in the network, will be able to perform various functions related to evidence. On performing any function, the transaction will be initiated and the mining of the transaction into the block will confirm the transaction. The evidence submitted or changed will be sent to IPFS, on sending the data, a hash will be generated. This ipfs hash will be stored in the blockchain with the wallet address that is logged in to the system.

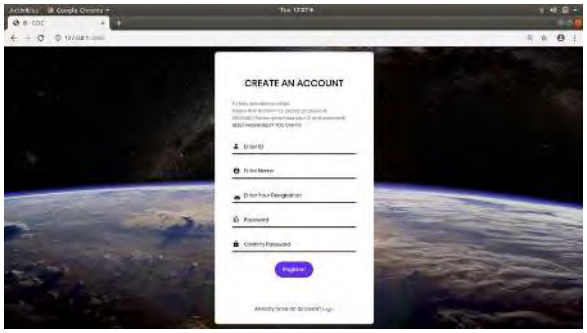
A. Methodology Used:

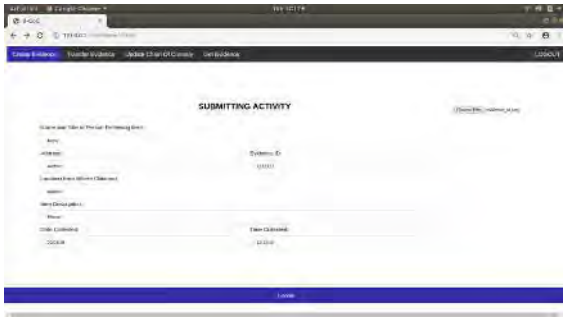
The proposed system will consist of a private Ethereum network connected to different nodes. The data will be stored on a decentralized file storage system (Interplanetary File System) and the hash of the evidence will be stored on the chain. It consists of python flask as a micro-framework. It will allow the user to create and upload the evidence and also transfer its ownership to another user in the network. The user will be the able to see the timeline of the evidence's custody in the update chain of custody module. These evidences will be stored in the IPFS and the hash associated with the evidences will be stored in the private Ethereum Blockchain. The evidences can be retrieved using the hash from get evidence module.

B. Algorithm:



C. User Interface Design(Snapshot):





IV. CONCLUSION

There are a lot of cases in our country for which justice is delayed. Judiciary is one of the strongest pillar in a democratic country. Our system is capable of reducing this delay to a considerable level as our current system is being exploited by many advocates and court officials. Current Indian Judicial System lacks accountability and transparency which will be provided by our system. Our system comprises of immutability for tamper-resistant legal record storage and management. Once uploaded into the network no one can modify, delete that evidence and once a new entry is added one cannot deny the existing ones. .

Our system has eliminated the paper based processes which will not lead to loss of evidences, duplication, tampering and errors. Our system will also assure that evidence once uploaded in private network will not be tampered or deleted. Our system makes the evidence traceable and secure. Our system will benefit the society at large level. Just as Bitcoin brought "banking for the unbanked", our system has the potential to bring "justness for the unjustness" using blockchain.

V. ACKNOWLEDGMENT

The success and final outcome of this project required guidance and assistance from many people and we are extremely privileged and honoured to have completed the project successfully.

We sincerely thank our Principal, Dr B. K. Mishra and HOD, Mr. Rajesh Bansode for always encouraging us to do our best. We are thankful to our guide Dr. Deven Shah who has supported and constantly supervised us throughout this project and helped us in completing this project and provided us with his unfathomable knowledge that was really beneficial to us.

We are thankful and fortunate enough to get constant

encouragement, support and guidance from all the Teaching staffs of IT Department who helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in laboratory for their orderly support.

We would like to express our gratitude towards our parents for their humble co-operation and encouragement which has helped us in completion of this wonderful project.

References

[1] *Kartik Hegadekatti . "Legal Systems and BlockchainInteractions", 2017.*

[2] *Dr. Robert Herian," Legal Recognition of BlockchainRegistries and Smart Contracts", 2018.*

[3] *Cl´ement Lesage and Federico Ast "Kleros- A Protocol of Decentralized Justice System", November 2018.*

[4] *Buterin, V. Ethereum, a next-generation smart contract and decentralized application platform. https://github.com/ethereum/wiki/wiki/White-Paper, 2014*

Theft detection using facial recognition via single-pixel detection with optical data security

Aman Nailwal, Umang Mishra, Mr.Vikas Kaul, Gopal Ojha
Information Technology
Thakur College of Engineering and Technology
Mumbai, India
amannailwal1@gmail.com, umangdmishra1@gmail.com, kaul.vikas@thaureducation.org, ojhagopal.98@gmail.com

Abstract— **An anti-theft system is any device or method used to prevent the unauthorized appropriation of items considered valuable. Surveillance video is used for watching known threats in real time. Video Surveillance is used to observe unusual activity going on to prevent crime."Theft detection using facial recognition via single-pixel detection with optical data security" is an Artificial Intelligence based surveillance and theft detection system which uses Al tools such as Machine Learning, Image Processing, facial recognition, etc without any human interaction involved such as security guards. The system works by placing a smart-phone in a private place as a security camera, the smart phone will be powered on and providing surveillance to the users which they can use via their android smart-phone. The system is loaded with pictures of genuine users who are not thieves, and when an unauthorized user (thief) tries to take away some belongings from there, the theft detection system immediately captures his picture and immediately notifies the user on their android smart-phone app. Hence the user is alerted and the identity of the thief is revealed and it is much easier to catch him then. Since this system does not involve any human interaction, hence it is highly cost efficient and prevents any possible human error. The system is an application that can run on Android operating systems which will do facial recognition via single pixel detection and will also provide optical data security. Basic GUI of the app is prepared using android studio to open front and rear camera of the smart phone to click pictures, and store it in mysql database these images are safe which are known to the user.**

Keywords—Android application, Facial recognition, Optical data security, pixel detection, AI based surveillance, smart phone, Theft detection.

I. INTRODUCTION

People around the world and specially in a country like India people are in a constant tautness of losing their Valuables to a potential theft. So addressing this issue we have come up with a system called as "Theft detection using facial recognition via single-pixel detection with optical data security"."Theft detection using facial recognition via single-pixel detection with optical data security" is an Artificial Intelligence based surveillance and theft detection system which uses Al tools such as Machine Learning, Image Processing, etc to solve one of the major problems of our country – Theft, without any human interaction involved such as security guards. The proposed system uses advance machine learning and image processing algorithms to detect theft and uses encryption to provide optical data security.

Video-based facial recognition is a promising technology that allows covert and unobtrusive monitoring of individuals[1]. Generally, video sequences are a collection of sequential static frames, thereby allowing the use of still-image-based techniques. However, in video-based techniques, one can utilize the temporal continuity of the image sequences to enhance robustness of the recognition process[2]. Single-pixel correlated imaging, widely known as ghost imaging, has been one of the most interesting topics in many application fields due to its remarkable characteristics. In single-pixel correlated imaging, the sample can be reconstructed by scanning the camera at reference beam arm, in which the sample is not located [2].

II. II.LITERATURE SURVEY

In paper [1] Single-pixel imaging via phase extraction is presented for optical security and flexibility enhancement. A series of random intensity patterns are pre-generated as principal security keys, and are sequentially encoded into phase-only masks. It can be advanced by using single-pixel correlated imaging as a quantum or classical technology with potential for significantly enriching the security field.

In paper [2] enhanced authentication method is used that uses video-based facial recognition technology to monitor the user during the entire session in which the person is using the system. Demonstration system uses facial recognition technology to periodically verify the identity of the user. If the authenticated user's face disappears, the system automatically performs a log-off or screen-lock operation. Uncontrolled face recognition from video can be done using newly developed algorithms to improve the facial presence monitoring system.

In paper [3] uses deep motion features as well as deep still-image features, following the success of two-stream convolutional networks, each of which are trained separately
for spatial and temporal streams. It is a Method for exploiting motion information in pedestrian detection with deep learning. The algorithms can be made more efficient so as to differ between human, still objects and animals effectively.

In paper [4] detection of events related to groups of people which are merge, split, walk, run, local dispersion, and evacuation by analysing the instantaneous optical flow vectors and comparing the learned models. It bypasses time-consuming methods such as background subtraction

and person detection and rather resorts to global motion information obtained from optical flow vectors to model the motion magnitude and velocity at each spatial location of the scene.

III. BACKGROUND

Artificial intelligence is a broad field which contains various subfields like image processing ,machine learning, deep learning, data science, NLP, etc. We planned to use the latest technologies and expertise using facial recognition with single pixel detection and optical data security in the best way to make a project that would try to solve one of the biggest security problems of the society and that is theft.

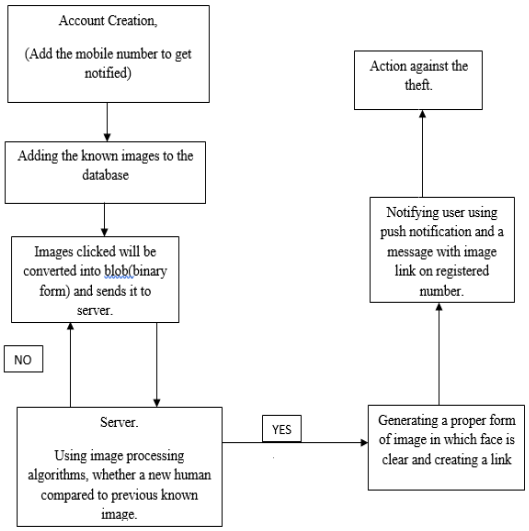
A study on image preprocessing algorithms has been carried out to improve the accuracy performance. It focused on the areas that affect the accuracy of facial recognition, which include geometric correction, face alignment, masking and photometric normalization. Single-pixel correlated imaging via phase extraction has been presented for optical security and flexibility enhancement. A series of random intensity patterns are pre-generated as principal security keys, and are sequentially encoded into phase-only masks.

IV. PROBLEM DEFINITION

We see a lot of theft going around the country like theft in the houses, apartments, jewelry shops, ATM machines, during train journey, etc. We need a system which will be smart and efficient so that the theft taking place around the country can be minimized to a great extent. The proposed system can detect theft and alert the user using advance machine learning and image processing algorithms through smart phone. Facial recognition through single pixel detection is used here with optical data security to provide data confidentiality.

V. METHODOLOGY

Traditionally, an authenticated user has to sit and look for any theft just like an video surveillance on any industry, also one of the upgraded application is in which an external security camera is placed to catch threat without any human interaction alerts the user about the threat there and thereby, whereas our system is just an application based on android operating system below is the detailed proposed architecture shown below.



VI. RESULTS & DISCUSSION

The app uses the camera of the phone to continuously take pictures of the person in front of it and processes it using image processing algorithms,it matches the picture with the already uploaded pictures of the family members and if image matches nothing happens but if it does not matches then an alert is sent to the user that a theft is being carried out. Basic GUI of the app is prepared using android studio to open front and rear camera of the smart phone to click pictures**,** and store it in mysql database these images are safe which are known to the user.We are using machine learning algorithm to generate a unique key for matching, image stored are in BLOB form which helps to protection against eavesdropping.

Facial detection is used to capture every pixels without any distortion in rgb values and image captured. Size of the image depends upon the pixels intensity and how many megapixel of camera is used hence to overcome it we will provide cloud api to enable unlimited cloud storage facility for seamless image processing, matching images, sending or receiving commands. Single-pixel imaging has a potential to prevent some attacks when the keys can be continuously modified. However, conventional single-pixel correlated imaging systems are usually established based on direct use or simple modulation of phase-only masks, and system flexibility is limited. It is also desirable that the indirect space for phase can be large to make single-pixel correlated imaging system difficult for unauthorized receivers. In this paper, single-pixel imaging via phase extraction is presented for optical security and flexibility enhancement. A series of random intensity patterns are pre-generated as principal security keys and are sequentially encoded into phase-only masks. Since different optical sensing infrastructures can be arbitrarily designed, respectively, for the extraction of

phase-only masks and object encoding, high flexibility and high security with a largely indirect (or virtual) space for phase are achieved in the proposed system.

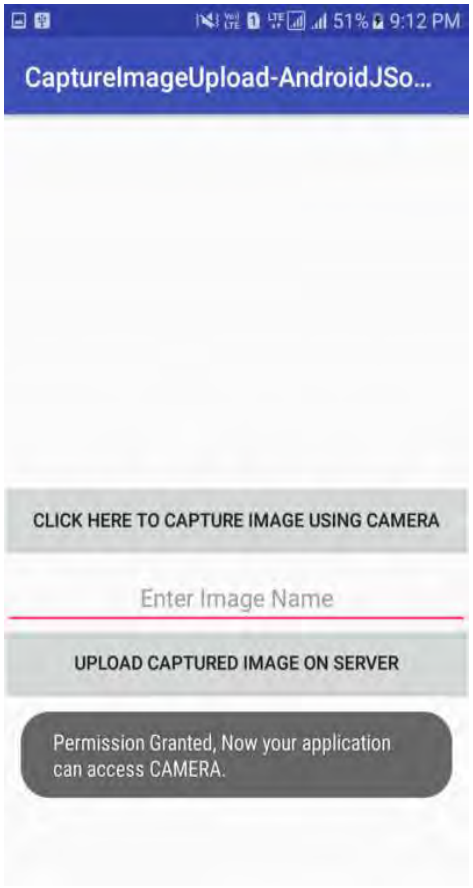


Fig:- 2 GUI to upload images on server.

Fig:- 3 GUI to activate Camera.

Fig:- 4 GUI to choose screen color and camera.

VII. CONCLUSION

This system will provide a completely automated theft detection and surveillance system by simply setting up a smartphone camera in their homes/offices/hotels etc. It will provide image recognition with optical data security. The app uses the camera of the phone to continuously take pictures of the person in front of it and processes it using image processing algorithms, it matches the picture with the already uploaded pictures of the family members and if image matches nothing happens but if it does not matches then an alert is sent to the user that a theft is being carried out.

Currently the application presents a GUI to add known images in the database either captured with camera or from an external storage, every image is converted into BLOB form. Lastly single-pixel detection is used to get the image in correct form with proper mathematical values which helps to generate unique key for matching it with known images, single-pixel imaging has a potential to prevent some attacks when the keys can be continuously modified. However, conventional single-pixel correlated imaging systems are usually established based on direct use or simple modulation of phase-only masks, and system flexibility is limited. It is also desirable that the indirect space for phase can be large to make single-pixel correlated imaging system difficult for unauthorized

receivers. In this paper, single-pixel imaging via phase extraction is presented for optical security and flexibility enhancement.

VIII. FUTURE SCOPE

Our current system is capable of identifying possible thefts along with capturing the criminal's face and reporting the owners. However the system can be further upgraded in the future to also report the incident to

the nearest police station and to get instant help from the police to catch the thief, we hope this move will highly discourage theft activities and will help the police catch the thieves swiftly at the time of crime itself. The system can be made more secure using better encryption techniques.

References

[1]. Wen Chen, "Optical Data Security System Using Phase ExtractionScheme via Single-Pixel Detection ",IEEE Photonics Journal, Volume 8, Issue 1, Feburary-2016.

[2]. Qinghan Xiao and Xue-Dong Yang, "Facial Recognition in Uncontrolled Conditions for Information Security EURASIP Journal on Advances in Signal Processing, Volume 2010, Issue 345743, Feburary2010.

[3]. Ryota Yoshihashi , Tu Tuan Trinh, Rei Kawakami, Shaodi You, Makoto Iida and Takeshi Naemura," Pedestrian detection with motion features via two-stream ConvNets ", YoshihashietalIPSJTransactionson Computer Vision and Application, Volume 10, Issue 12, 2018.

[4]. Yassine Benabbas, Nacim Ihaddadene, and Chaabane Djeraba, "Motion Pattern Extraction and Event Detection for Automatic Visual Surveillance", EURASIP Journal on Image and Video Processing, Volume 2011, Issue - 163682,2010.

[5]. Alexia Briassouli, Vagia Tsiminaki, and Ioannis Kompatsiaris, "Human Motion Analysis via Statistical Motion Processing and Sequential Change Detection", EURASIP Journal on Image and Video Processing, Volume 2009,Issue 652050,2009.

[6]. Mohd Rizal Mohd Isa , Salem Aljareh1 ,Zaharin Yusoff, "A watermarking technique to improve the security level in face recognition systems ", Multimed Tools Appl , Volume 76, Issue 23805–23833,November-2016.

[7]. Zaid Al-Ars , Steven van der Vlugt, Pekka J ainen , Frank van derLinden, "ALMARVI System Solution for Image and Video Processing in Healthcare, Surveillance and Mobile Applications", Journal of Signal Processing Systems, Volume 91, Issue 1-7, November-2018

Data Security using Image Encryption Algorithm

Ritesh Gupta[1], Kaushal Solanki[2], Nikhil Gupta[3], Rahul Neve[4]
Department of Information Technology, Thakur College of Engineering and Technology
Kandivali (East), Mumbai-400101, India
guptaritesh2818@gmail.com[1], kaushalsolanki1999@gmail.com[2], nikhildgupta@gmail.com[3], rahulneve@gmail.com[4]

Abstract— **Security is that the major issue in today's world as several digital services, like Medical, Military, and house imaging systems need reliable security in storage and transmission of digital pictures. The speedy progress of net within the digital world these days, the protection of digital pictures has become additional and more vital to meet such security and privacy desires in numerous applications, secret writing of pictures is extremely vital to attenuate malicious attacks from unauthorized parties. This proposed system is meant to the people that would like privacy for his or her confidential pictures. We will see however antecedently planned strategies like encryption customary (DES), Triple encryption customary (Triple-DES), and International encryption algorithmic rule (IDEA) are applied in image protection domain and the way new ideas of Chaos-based encryption techniques are superior to ancient strategies. Our proposed system provides an answer to existing system by extending its facilities. The planned study aims to explore the likelihood of victimization chaotic or chaos-based secret writing techniques to guard essential pictures and provides high level of security in economical and reliable means.**

Keywords—Security, Attacks, Digital Image Encryption, Chaos-based encryption

I. INTRODUCTION

The development of the area Science and Technology has recently attracted a growing interest from researchers and industrial communities, in the main thanks to sizable amount of doable applications capable to take advantage of remotely perceived information and satellite pictures. Advances in area science, information analysis, and communication technologies gift new opportunities for users to extend productivity, cut back prices, facilitate innovation and make virtual cooperative environments for addressing the new challenges.

The speedy progress of web within the digital world nowadays, the protection of digital pictures has become additional and more necessary .The commonness of mixed media framework innovation in our general public has elevated advanced pictures to assume an extra essential job that requests a critical security of users‟ protection. To fulfill such security and protection needs in various applications, mystery composing of pictures is incredibly important to lessen pernicious assaults from unapproved parties. Digital image encryption is one of the safest methods of securing digital images when transmitted over unse cure channels from unauthorized copying.Sadly, acco rding to [1][2][3] several famous methods of encrypt ion like DES, RSA, AES and others function only we ll for plaintext but not for digitalimages. In most natur al digital images, the values of the neighboring pixels are closely corelated.

This unusual feature will lead to tremendous changes in each pixel of the plain image will not significantly reduce the quality of the cipher image which will mak e the content of the cipher image still easily recogniza ble by humans.

One way to solve these problems is to use chaotic sys tem (i.e. chaotic logistic map) in a cipher since, accor ding to [4][5], chaos is very sensitive to a small chan ge in the initial value and will generate the same effec t as diffusion and uncertainty.

II. MOTIVATION

The total software development life cycle (SDLC) inv olves multiple actors throughout the typical applicatio n creation

and maintenance approach.In terms of features, expen se, schedule, quality, efficiency, and other administrat ive overheads such as vendor integration, this approac h is often granulated by strategies, systems, processes , staff, and resources that

Impact the final product.Throughout the project lifecy cle, the SDLC operations and production engineer's work together from design through to the developmen t process to assist in the entire production process.

Security is been our most important aspect has most of the people are going online so for their protection we needed to do something.

III. BACKGROUND

There are many data hiding techniques some turn the data into non-understandable format, whereas some hide the existence of secret data. Few of the data hiding mechanisms are Cryptography, Steganography and Digital Signature.

1. Cryptography

The Cryptography involves 2 major processes – cryptography and coding. The method of conversion of plain text to cipher text is understood as cryptography whereas the process of conversion of cipher text back to plain text is known as coding.

In cryptography, users (sender and also the receiver) select associate formula to be adopted. This formula is then accustomed first encipher the key knowledge at the sender's finish so followed by the recipient to rewrite the encrypted data so as to reveal the initial secret message. For example: Ceaser Cipher formula.

2. Steganography

Steganography is that the art of activity information in such the way that the existence of the key data is understood solely the concerned parties.

In steganography, data is hidden during a medium referred to as cowl medium in such the way that nobody with the exception of the sender and therefore supposed recipient are aware of it. So, it doesn't attract unwanted attention of the attackers.

3. Digital Signature

A propelled imprint is a logical arrangement for displaying the realness of an automated message or record. A generous automated mark gives a recipient inspiration to acknowledge that the message was made by a known sender, with the end goal that the sender can't deny having sent the message and that the message was not changed in travel. Computerized marks are normally utilized for programming dissemination, budgetary exchanges, and in different situations where it is imperative to recognize phony or altering. Two principle properties are required. Initially, the credibility of a mark created from a fixed message and fixed private key can be checked by utilizing the relating open key. Also, it ought to be computationally in practical to create a substantial mark for a gathering without realizing that gathering's private key. An advanced mark is a verification instrument that empowers the maker of message to append a code that go about as a mark. It is shaped by taking the hash of message and encoding the message with maker's private key. A digital signature scheme typically consists of three algorithms:

1. A key age calculation that chooses a private key consistently at arbitrary from a lot of conceivable private keys. The calculation yields the private key and a comparing open key.
2. A marking calculation that, given a message and a private key, delivers a mark.
3. A signature checking calculation that, given a message, open key and a mark, either acknowledges or dismisses the message's case to validness.

IV. PROBLEM DEFINITION

Data sharing is regularly established on CD/DVD-ROM printed duplicate or on shared framework condition (Internet, LAN, WAN, etc), so there exists common security threat of unapproved access or usage of the thing. To fulfill such security and insurance needs in various applications, encryption of such data is fundamental to constrain attacks from unapproved parties and to protect unstable data.

Stage 1: Image gets gathered from our framework with appropriate configuration for encryption

Stage 2: The picture is then stacked in the product which will get scrambled utilizing the encryption calculation

Stage 3: The encoded picture is then send over system which gets by the client to get decoded utilizing the key gave to him.

V. OBJECTIVE

Despite the fact that there are many picture encryption strategies none of them are reasonable for the systems administration frameworks. So the primary extent of our venture was to give security to the pictures in the systems administration systems. Our venture gives safe methods for intends to move pictures between the systems administration frameworks privately.

This undertaking is planned to the individuals who need security for their classified picture It was generally reasonable in the systems administration frameworks, so it was increasingly qualified in Space Science look into focuses.

VI. LITERATURE SURVEY

We have undergone certain research papers based on the Image Encryption Algorithm and found out certain methodology and key findings. During the process, we have also identified certain research gaps which we can overcome while implementing our project

Venkata Krishna Pavan Kalubandi, Hemanth Vaddi Vishnu, Ramineni,Agilaeeswari Loganathan:

A Novel Image Encryption Algorithm using AES and Visual Cryptography

The cryptographic methodology proposed in this paper has been tested on different types of input images with change in size of the image and keys of AES encryptionThe confidentiality of shares is also tested by changing the key shares before reaching to the destination.

This scheme can possibly be modified to use color image in place of binary image for the key and then generate the shares using Visual Cryptography.

Bhaskar Mondal a, Tarni Mandal:

A light weight secure image encryption scheme based on chaos & DNA computing

The proposed algorithms use logistic map and DNA sequencing and is used in the substitution phase of the encryption process which makes it light weight and resistant against statistical attacks. Permutation is also done on a plain image which gives better performance and quality.

The scheme is proposed for gray label images but the scheme may be extended for color images and text data.

Riah Ukur Ginting,Rocky Yefrenes Dillak:

Digital Color Image Encryption Using RC4 Stream Cipher and Chaotic Logistic Map

Experimental results show that our method can be used as an alternative method to encrypt digital images because this method: (i) can encrypt image in such way so that cipher-image cannot be visually identified by human, (ii) eliminates statistical relation between plain-image and cipher-image.

Since in this algorithm the key stream doesn't rely on plain text in general .Due to its simplicity, vernam stream cipher is widely used as it's depends on variable key-size.

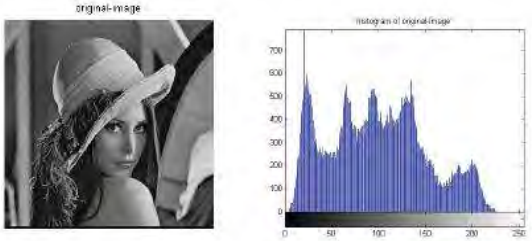
Kamlesh Gupta, Sanjay Silakari, Ranu Gupta , Suhel A. Khan:
An Ethical way for Image Encryption using ECC

The resulting system gives comparatively small block size, high speed and high security. The synthesized texture by an Mrf using the authorized key derived from elliptic curve cryptography is presumably difficult to be revealed.
The proposed system is bounded with an using Mrf texture model to cover a secret message to achieve information concealment before doing a conventional RSA data encryption.

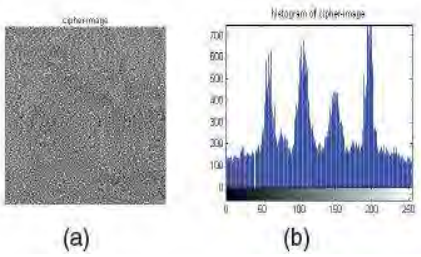
Minal Govind Avasare,Vishakha Vivek Kelkar:
Image Encryption using Chaos Theory
Relations between chaos and encryption are: a. Output has a distribution similar, b. Output changes caused by input , c. System for generating pseudorandom signal.
By doing PSNR and computational time analysis we get that above algorithm are weaker than chaos based techniques.

VII. ANALYSIS

Experimental analysis:
This is the software which has brought security in the image
Form and provide the ability to encrypt it using



Original image to histogram image



Chiper image

Security analysis:
The hidden keys should be immune to a good encrypt ion scheme, and the key space should be large enough to make Bruteforce attacks infeasible.The keyimage and initial valus of chaotic map are used in our encryption algo rithm as hidden keys.The main space is wide enough t o withstand
Bruteforce attacks of all kinds. The experimental resu lts also indicate that our scheme is very sensitive to th e mismatch of the main secrets.

Cipher	Key length	Key space
Vigenère	128	2^{128}
RC4	256	2^{256}
Salsa20	256	2^{320}
DES	56	256
3DES	168	2^{168}
AES	128	2^{128}
ECC	2*64	2^{128}
Chaotic map[49]	3*50	2^{150}

Fig; Analysis of different algorithms key space.

VIII. METHODOLOGY

In our Image Encryption tool, we have a login page in which users will login with their password. After sign in on the tool, login details are checked and if correct we insert the image on the tool on which the encryption mechanism is applied and then image is saved .After that the key is generated and the encrypted image is send across to the network to the receiver. After receiving the image at the receiver's end, the image is decrypted by using the key which was generated during encryption and then the original image is obtained.

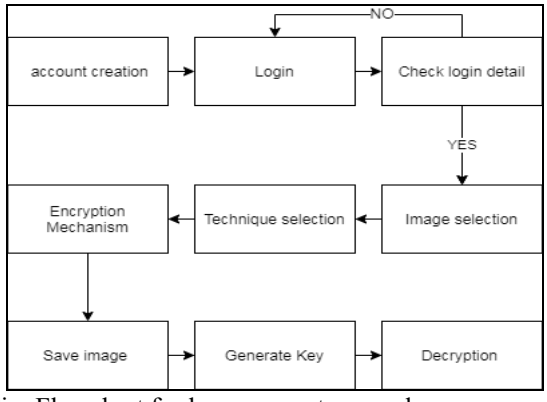


Fig: Flowchart for how our system works.

IX. RCHITECTURE

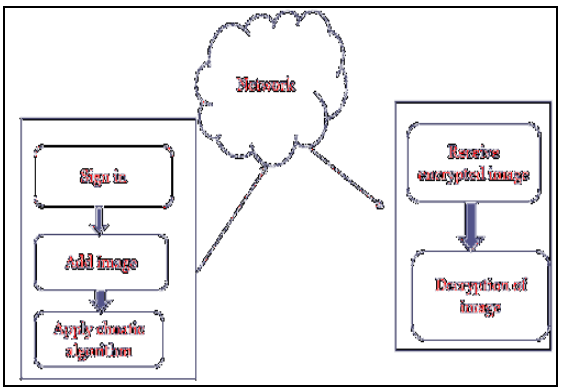


Fig. User interface contains the sign in module which consist of adding an image and applying chaotic algorithm. The encrypted image get transferred over the network to the receiver. The receiver then try to decrypt the encrypted image by using the subsequent key

Concerning the administrator the user can have memb ership. The system generates member ids dynamicall y when presenting user details. Through getting your member I d the user can have transactions. The user a ttaches the encryption file, and encrypts it for sensitiv e images.

Users of the system:

Administrator:

Administrator has all the access rights to this site. Admin will display all of the files, and the encryption and decryption process status. He will have network power over whether to approve the request for image receipt from the server. He will provide all of the image encryption and decryption benefits.

Users:

Users are the people who use this program to protect sensitive photos. This program may have different set of users. The users just load an image to encrypt and encrypt it and enter the destination systems IP address to transfer the image and he can view the status of the network for the sent image.

Guest:

Only the functionalities in the portal can be accessed by guest or new user, and they have no access rights unless approved.

X. FUTURE SCOPE

This venture speaks to a starter investigation of various components utilized for picture security featuring far reaching relative diagram of existing customary picture encryption methods like DES, Triple-DES and IDEA calculations. Some present commotion based picture encryption plans have also been immediately assessed by mulling over and inspecting various computations properties, for instance, encryption speed, comparability to picture association and weight rules, and continuous execution, etc. Presently a day‟s tumult based encryption system is getting increasingly more fame worldwide to manage interactive media informational collections particularly for pictures and is prescribed by numerous specialists. The equivalent methodology of uproar based picture encryption is needed to be applied on satellite imageries since disturbance based strategies have various genuine associations with achieve high security level in gainful and reliable manner, like affectability, ergodicity, intervention, and no normal or returning circles for huge stretches with high spread and chaos. The outline of the planed future work is the accompanying: Continue to think about other proposed encryption methods, hypotheses and calculations, for example, tumult based or disorderly cryptosystem, number hypothesis, AES, DES, RSA and so forth., should be comprehended. Research and analyze the security and heartiness of some disarray based and conventional encryption plans for huge informational indexes for example gigantic satellite pictures. Inside and out examination general frailty properties of satellite symbolism and remotely detected information, and attempt to propose some relating countermeasures, moreover, structure some new plans that have great security properties and can meet the genuine application prerequisites simultaneously.

XI. CONCLUSION

Advances in space science, information appraisal, and correspondence advances present new open passages for clients to expand profit, lessen cost, engage improvement and make virtual system masterminded conditions for watching out for new difficulties. In such procedures information sharing is all around built up on CD/DVD-ROM printed adjustment or on shared system condition (Internet, LAN, WAN, and so on) , so there exists regular security danger of unapproved access or utilization of the thing. To satisfy such security and confirmation needs in different applications, encryption of such information is fundamental to oblige malignant ambushes from unapproved parties and to shield delicate information. From the assessment of the above conventional and dissipate based picture encryption systems; it is ended clear that standard picture encryption methods DES, Triple-DES and IDEA have two or three impediments, for example, these figuring's have high security level under CBC mode yet requires enormous information size, long computational time and high getting ready power.

XII. ACKNOWLEDGEMENT

The quality and ultimate outcome of this paper involved a great deal of support and expertise from many people and

we are incredibly privileged to have successfully completed the paper. We take this opportunity to thank everybody

for their support. We sincerely thank our principal Dr. B. K. Mishra, vice Principal, Dr. Deven Shah, Mentor Dean Dr. Kamal Shah and HOD Dr. Rajesh Bansode for always guiding us to paths that were right for us. We are grateful to our guide Mr. Rahul Neve who supported and constantly monitored this work and helped us in not only completing this project but also

providing us with basic knowledge with their practical aspect that really proved beneficial to us. We are forever thankful and fortunate enough to receive constant encouragement, support and guidance from all IT department faculties that helped us finalize our paper work successfully. We would also like to extend our sincere gratitude to all the laboratory support staff for their timely support. We would like to convey our heartfelt thanks to our parents for their kind cooperation and encouragement without which none of this was possible.

References

[1]. Venkata Krishna Pavan Kalubandi,Hemanth Vaddi, Vishnu Ramineni, Agilaeeswari Loganatha**.,** **"**A Novel Image Encryption Algorithm using AES and Visual Cryptography**"** 2nd International Conference on Next Generation Computing Technologies (NGCT-2016)Dehradun, India 14-16 October 2016

[2]. Bhaskar Mondal a,, Tarni Manda**.,** **"**A light weight secure image encryption scheme based on chaos & DNA computing**"** Department of Computer Science and Engineering, National Institute of Technology Jamshedpur, Jamshedpur, Jharkhand 831014, India

[3]. Riah Ukur Ginting,Rocky Yefrenes Dillak**.,** **"**DIGITAL COLOR IMAGE ENCRYPTION USING RC4 STREAM CIPHER AND CHAOTIC LOGISTIC MAP**"** ,978-1-4799-0425-9/13/2013 IEEE

[4]. Kamlesh Gupta, Sanjay Silakari, Ranu Gupta , Suhel A. Khan**.,** **"**An Ethical way for Image Encryption using ECC**"** 2009 First International Conference on Computational Intelligence, Communication Systems and Networks.

[5]. Minal Govind Avasare,Vishakha Vivek Kelkar., **"**Image Encryption using Chaos Theory**"** 2015 International Conference on Communication, Information & Computing Technology

[6]. Ray A. Williamson, "REMOTE SENSING AND TRANSPORTATION SECURITY",Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS Conference Proceedings, 2002 (ICCICT), Jan. 16-17, Mumbai, India

[7]. Pareek N.I<, Patidar V, and Sud K.K, 2006. Image encryption us in chaoticlogistic map, Image and Vision Computing 24 926-934.

[8]. Mazloom, S., Eftekhari-Moghadam, A., 20 11, Color Image Cryptosystem using Chaotic Maps, IEEE, 978-1-4244-99 15-1.

[9]. Liu Xiangdong, Zhang Junxing, Zhang Jinhai, He Xiqin, "Image scrambling algorithm based on chaos theory and sorting transformation",International journal of computer science and network security, Vo1.8No. 1, January 2008.

[10]. Schneier B. Applied Cryptography, Second Edition, New York: John Wiley & Sons, Inc. 1996.

URL Anti Phishing using Heuristic Approach

Abhishek Jain, Jinesh Khatiya, Vedant Mandre, Dr. Sangeeta Vhatkar
Information technology, Thakur College of Engineering and Technology, Mumbai india
hashabhishekjain@gmail.com, jineshkhatiya@gmail.com, vedantmandre29498@gmail.com, Vhatkarsangeeta@gmail.com

Abstract—**Damage caused by phishing attacks that target personal user information is increasing. Phishing involves sending an email to a user or inducing a phishing page to steal a user's personal information. This type of attack can be detected by blacklist-based detection techniques; however, these methods have some disadvantages and the numbers of victims have therefore continued to increase. In this paper, we propose a heuristic-based phishing detection technique that uses uniform resource locator (URL) features. We identified features that phishing site URLs contain. The proposed method employs those features for phishing detection. The technique was evaluated with a dataset of 3,000 phishing site URLs and 3,000 legitimate site URLs. The results demonstrate that the proposed technique can detect more than 98.23% of phishing sites.**

Keywords—phishing sites, URL-based features, heuristic, machine learning

I. INTRODUCTION

With the recent growth of the Internet environment and diversification of available web services, web attacks have increased in quantity and advanced in quality. Phishing is a type of social engineering attack that targets a user's sensitive information through a phony website that appears similar to a legitimate site, or by sending a phishing email [1]. According to research of the Anti Phishing Working Group (APWG), 85,062 phishing sites were globally detected in the second quarter of 2010; by the second quarter of 2014, 128,978 were detected. These figures mark an increase of 1.5 times the value that count of occurred phishing attack in one quarter [2,3]. In addition, annual damage caused by phishing was measured at $5.9 billion. Thus, phishing is a worldwide malicious activity that continues to increase.

In response to this increase in phishing attacks, phishing detection techniques have been the focus of considerable research. Typical phishing detection techniques include the blacklist-based detection method and the heuristic-based technique. The blacklist-based technique maintains a uniform resource locator (URL) list of sites that are classified as phishing sites; if a page requested by a user is present in that list, the connection is blocked [4]. This technique is commonly used and has a low false-positive rate; however, its accuracy is determined by the quality of the list that is maintained. Consequently, it has the disadvantage of being unable to detect temporary phishing sites [5].

Abhishek Jain, Jinesh Khatiya,
Vedant Mandre
Mumbai University

** This research was supported by Thakur College of Engineering and Technology **

The heuristic-based detection technique analyzes and extracts phishing site features and detects phishing sites using that information [6]. In this paper, we propose a new heuristic-based phishing detection technique that resolves the limitation of the blacklist-based technique. We implemented the proposed technique and conducted an experimental performance evaluation. The proposed technique extracts features in URLs of user-requested pages and applies those features to determine whether a requested site is a phishing site. This technique can detect phishing sites that cannot be detected by blacklist-based techniques; therefore, it can help reduce damage caused by phishing attacks.

II. RELATED WORKS

Phishing is an attempt to steal a user's personal information typically through a fraudulent email or website [1]. We conducted a study on phishing sites, which are either fake sites that are designed to appear similar to legitimate sites or sites that simply have phishing-related behaviors. Almost all phishing sites include the functionality in which users enter sensitive information, such as their personal identification, password, and/or account number. These sites can include links to connect to other phishing sites and malicious code that contaminates a user's computer.

Phishing detection techniques can be generally divided into blacklist-based and heuristic-based approaches. The blacklist-based approach maintains a database list of addresses (URLs) of sites that are classified as malicious. If a user requests a site that is included in this list, the connection is blocked [4]. The blacklist-based approach has the advantages of easy implementation and a low false- positive rate; however, it cannot detect phishing sites that are not listed in the database, including temporarily sites [5].

The heuristic-based approach analyzes phishing site features and generates a classifier using those features [6]. When a user requests a web page, the classifier determines whether that page is a phishing

site. This approach can detect new phishing sites and temporary phishing sites because it extracts features from the requested web page. Nevertheless, it has the disadvantage of being difficult to implement; moreover, generating a classifier is time- intensive. Thus, the two approaches have both advantages and disadvantages. Therefore, these approaches are selectively employed in the proposed technique depending on the application.

III. PROPOSED APPROACH

A. URL Structure

A URL is a protocol that is used to indicate the location of data on a network. The URL is composed of the protocol,

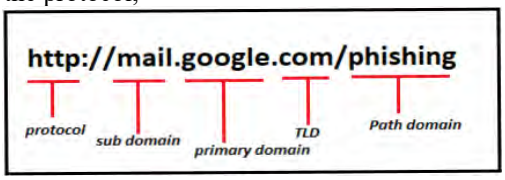


Figure 1. URL Structure.

subdomain, primary domain, top-level domain (TLD), and path domain. [6]. In this study, the subdomain, primary domain, and TLD are collectively referred to as the domain. Fig. 1 depicts the individual components of a URL.

The protocol refers to a communication protocol for exchanging information between information devices; e.g., HTTP, FTP, HTTPS, etc. Protocols are of various types and are used in accordance with the desired communication method.

The subdomain is an ancillary domain given to the domain and has various types depending on the services provided by the domain page. The domain is the name given to the real Internet Protocol (IP) address through the Domain Name System (DNS). The primary domain is the most important part of a domain. The TLD is the domain in the highest position in the domain name hierarchy architecture; e.g., .com, .net, .kr, .jp, etc. [7]. We define features of each component of the URL; these features are used for phishing site detection.

B. URL Features

Table 1 shows 26 URL-based features that are used in the proposed detection technique.

- Features 1 to 6 relate to Google Suggestion. They return a suggested word when a user enters a single term. We analyze the results of Google Suggestion when entering the URLs of phishing sites and legitimate sites. If a search term is similar to a suggested result, input URL is doubtful because that site may be emulating an existing site. We use Levenshtein distance between the two terms—the Google Suggestion result and the search term—as a feature for detecting phishing sites [6,8]. In addition, if a suggested result is the same as that of a domain that is present in the trustworthy whitelist, that search term site may be emulating a legitimate site [8]. For this reason, we can detect phishing sites using this feature.

- Features 7 to 9 can be extracted through page ranking. The page rank is a numerical value that is calculated by the number of visitors and degree of popularity. Phishing sites have a very low page rank value or no value because phishing sites are not often visited by many people and they exist for a short time [9]. Therefore, if a domain page rank value is very low, it can be regarded as a phishing site.

- Features 10 to 14, and 16, 17, 19, and 20, are associated with suspicious URL patterns and characters. Characters such as _@‘ and _//‘ rarely appear in a URL. Moreover, URLs of legitimate sites typically have one TLD. Therefore, patterns of many TLDs in a URL signify a fraudulent site [10]. Therefore, in the above cases, we classify these sites as phishing sites.

- Features 21 to 26 are characterized by URL property values. Because temporary phishing sites, as mentioned, often do not contain the required property values [11], property values can be used as features for identifying phishing sites.

Thus, many URL features exist that have been employed in several studies on phishing detection. In the present research, we incorporate features used in previous studies and define two new features for identifying existing phishing sites.

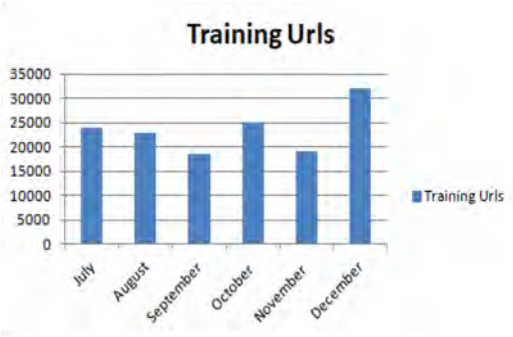


Figure2. Representation of total urls tested in a month

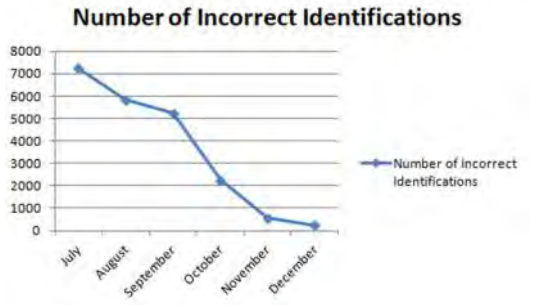


Figure3. Representation of total number of urls that are incorrectly

identified as phished website.

- Feature 15 is defined for identifying newly created phishing sites with the proposed technique. Currently, to prevent a user from recognizing that a site is not legitimate, phishing sites typically hide the primary domain; the URLs of these phishing sites have unusually long subdomains. Therefore, we added a feature that calculates the subdomain length of a URL to determine if it is a phishing site. This feature can be additionally used to identify phishing sites that target vulnerabilities of smartphones, which have small displays that make it difficult to see the full URL.
- Feature 18 is another new feature that reflects current phishing trends. This feature includes eight words that are predefined as phishing terms. It is verified if a requested page's URL contains these phishing terms [12][1]. This feature worked well in previous studies; however, we determined that changes have occurred since the studies were conducted. Thus, through experiments, we identified eight new phishing terms[2] and we employ them in our phishing detection technique.

As noted above, our proposed method employs new features that have not been previously used in studies. IN addition, it advances features from previous works to provide better phishing detection performance.

C. Architecture

The given below figure illustrates the proposed phishing detection process, which includes two phases: training and detection.

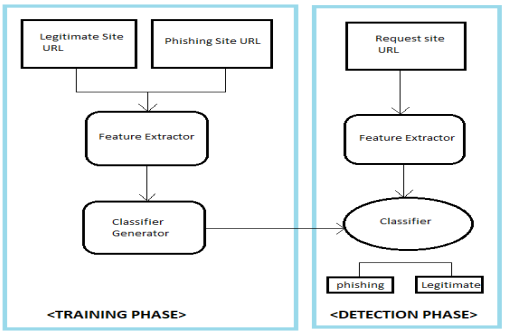


Figure 4. Process of proposed detection method.

In the training phase, a classifier is generated using URLs of phishing sites and legitimate sites collected in

advance. The collected URLs are transmitted to the feature extractor, which extracts feature values through the predefined URL-based features. The extracted features are stored as input and passed to the classifier generator, which generates a classifier by using the input features and the machine learning algorithm.

In the detection phase, the classifier determines whether a requested site is a phishing site. When a page request occurs, the URL of the requested site is transmitted to the feature extractor, which extracts the feature values through the predefined URL-based features. Those feature values are inputted to the classifier. The classifier determines whether a new site is a phishing site based on learned information. It then alerts the page-requesting user about the classification result.

D. Survey

To determine a classifier with the best performance for using URL-based features, we employed several machine learning algorithms: support vector machine (SVM), naive Bayes, decision tree, k-nearest neighbor (KNN), random tree, and random forest.

- SVM is a classification method that was introduced in 1992 by Boser, Gyon, and Vapnik [13]. It is a statistical learning algorithm that classifies samples using a subset of the training samples, called support vectors. SVM is built on the structural risk minimization principle for seeking a decision surface that can separate data points into two classes with a minimal margin between them [14]. The advantage of SVM is its capability of learning in spare high-dimensional spaces with very few training samples.
- Decision tree is a classification method that was introduced in 1992 by Quinlan [15]. It

creates a tree form for classifying samples. Each internal node of the tree corresponds to a feature, and the edges from the node separate the data based on the value of the feature [15]. Decision tree includes a decision area and leaf node. The decision area checks the condition of the samples and separates them into each leaf node or the next decision area. The decision tree is very fast and easy to implement; however, it has the risk of overfitting.

- Naive Bayes is a classifier that can achieve relatively good performance on classification tasks. It is based on the elementary Bayes‘ theorem. It is one of most successful learning algorithms for text categorization [16]. On account of the conditional model‘s feature, naive Bayes is effectively trained in supervised learning. It provides the advantage of learning essential parameters using small training samples.
- KNN is a non-parametric classification algorithm [17]. It has been successfully applied to various information-retrieval problems. It classifies the input data using k training data that is similar to the input data. KNN uses Euclidean distance to calculate the similarity between the input and training samples. Its performance is determined bythe choice of k; nevertheless, choosing a suitable k
value is not easy.
- Random tree is a tree-based classification method that was introduced by Breiman and Cutle [18]. A tree is drawn at random from a set of possible trees.
―At random‖ means that each tree in the set of trees has an equal chance of being sampled [18]. The random tree classifier takes as input a features vector and classifies it with each tree. The output is determined by the majority ―vote. This algorithm can handle both classification and regression.
- Random forest is a classification method that combines many tree predictors; each tree depends on the values of a random vector that is independently sampled [19]. All trees in the forest have the same distribution. This algorithm can handle a large number of variables in the dataset; however, it lacks reproducibility because the process of forest building is random [20].

IV. EVALUATION

To conduct classifier training and evaluation through an experimental dataset, we collected the URLs of phishing and legitimate sites. We gathered 3,000 phishing site URLs from PhishTank and 3,000 legitimate site URLs from DMOZ. The evaluation was conducted using k-fold cross validation. K-fold cross validation divides the input data into k; k – 1 datasets are used for training, and the remaining one is used for validation. This process is performed k times, such as the number in the divided dataset, because all datasets can be used for training and validation. This method is typically used to evaluate the accuracy of the classifier with a small dataset. In this study, we used ten-fold cross validation to evaluate our detection technique. We performed the testing with the WEKA open-source machine learning tool, and we analyzed the performance of each of the machine learning algorithms noted in Section 3. The accuracy was calculated as TP (true positive), TN (true negative), FP (false positive), and FN (false negative). We compared the performance of each classifier using the calculated accuracy. Fig. 3 depicts the TP, TN, FP, and FN matrix.

Figure 5. TP, TN, FP, FN matrix

TP is the ratio of the prediction that a determined phishing site is indeed a phishing site, and FN is the ratio of the prediction that a determined phishing site is actually a legitimate site. In addition, FP is the ratio of the prediction that a truly legitimate site is a phishing site, and TN is the ratio of prediction that a determined legitimate site is indeed a legitimate site. Table 2 shows the TP, TN, FP, FN ratios of each machine learning algorithm.

As a result of the experiments, we obtained TP, TN, FP, and FN ratios to calculate three measurements that we used to compare the performance of each algorithm. The first

measurement was for the specificity of the true negative rate. The second was the sensitivity of the true positive rate. The third was the accuracy of the total ratio of the prediction that a determined phishing site is actually a phishing site, and that a determined legitimate site is indeed legitimate.

TABLE I. TF, TN, FP, AND FN OF MACHINE LEARNING ALGORITHMS

Algorithm	Measurements			
	TP	TN	FP	FN
SVM	97.00%	94.90%	5.10%	3.00%
Decision Tree	96.90%	96.90%	3.10%	3.10%
Naive Bayes	90.90%	95.10%	4.90%	9.10%
KNN (k = 1)	96.30%	96.00%	4.00%	3.70%
Random Tree	96.10%	96.00%	4.00%	3.90%
Random Forest	98.10%	98.40%	1.60%	1.90%

In measuring the classifier performance, (1)

was the equation of specificity, (2) was the equation of sensitivity, and (3) was the equation of accuracy.

		Prediction	
		Positive	Negative
Actual	True	*True Positive*	*False Negative*
	False	*False Positive*	*True Negative*

$$\text{Specificity} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Sensitivity} = \frac{FP}{FP+FN} \quad (2)$$

$$\text{Accuracy} = \frac{TP+TN+FP+FN}{TP+TN} \quad (3)$$

We additionally used the specificity, sensitivity, and accuracy as measurements of the classifier performance measurement. Table 3 shows the specificity, sensitivity, and accuracy of each machine learning algorithm that we used in training.

TABLE II.EXPERIMENTAL RESULTS OF ALGORITHMS

Algorithm	**Measurements**		
	Specificity	**Sensitivity**	**Accuracy**
SVM	96.93%	95.00%	95.95%
Decision Tree	96.90%	96.90%	96.88%
Naive Bayes	91.26%	94.88%	93.01%
KNN (k = 1)	96.28%	96.01%	96.18%
Random Tree	96.09%	96.00%	96.03%
Random Forest	98.10%	98.30%	98.23%

As a result of the experiment, we determined that the best machine learning algorithm, random forest, used URL features. This classifier detected more than 98.23% of phishing sites. The high accuracy shown in Table 3 and low false-positive rate shown in Table 2 meant that the proposed phishing detection technique can effectively classify sites as either being phishing or legitimate..

V. CONCLUSION

In this paper, we proposed a heuristic-based phishing detection technique that employs URL-based features. The method combines URL-based features used in previous studies with new features by analyzing phishing site URLs. Additionally, we generated classifiers through several machine learning algorithms and determined that the best classifier was random forest. It showed a high accuracy of 98.23% and a low false-positive rate. The proposed technique can provide security for personal information and reduce damage caused by phishing attacks because it can detect new and temporary phishing sites that evade existing phishing detection techniques, such as the blacklist-based technique.

In future work, we intend to address the time-intensive disadvantage of the heuristic-based technique. With a large number of features, it is time-consuming for the heuristic- based approach to generate classifiers and perform classification. Therefore, we will apply algorithms to reduce the number of features and thereby improve performance. In addition, we will examine a new phishing detection technique that uses not only URL-based features, but also HTML and JavaScript features of web pages to improve performance

References

[1] Khonji, Mahmoud, Youssef Iraqi, and Andrew Jones. "Phishing detection: a literature survey." Communications Surveys & Tutorials, IEEE 15.4 (2013): 2091-2121.

[2] Anti Phishing Working Group. (2015. March.) APWG Phishing Activity Trend Report 2nd Quarter 2010. [Online]. Available: http://docs.apwg.org/reports/apwg_trends_report_q2_2014.pdf

[3] Anti Phishing Working Group. (2015. March.) APWG Phishing Activity Trend Report 2nd Quarter 2014. [Online]. Available: http://docs.apwg.org/reports/apwg_report_q2_2010.pdf

[4] Huang, Huajun, Junshan Tan, and Lingxi Liu. "Countermeasure techniques for deceptive phishing attack." New Trends in Information and Service Science, 2009. NISS'09. International Conference on. IEEE, 2009.

[5] Ma, Justin, et al. "Beyond blacklists: learning to detect malicious web sites from suspicious URLs." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009

[6] Nguyen, Luong Anh Tuan, et al. "A novel approach for phishing detection using URL-based heuristic." Computing, Management and Telecommunications (ComManTel), 2014 International Conference on. IEEE, 2014.

[7] Wikipedia. (2015. March) Uniform Resource Loactor. Avaliable: http://en.wikipedia.org/wiki/Uniform_resource_locator

[8] Kausar, Firdous, et al. "Hybrid Client Side Phishing Websites Detection Approach." International Journal of Advanced Computer Science and Applications (IJACSA) 5.7 (2014).

[9] Sunil, A. Naga Venkata, and Anjali Sardana. "A pagerank based detection technique for phishing web sites." Computers & Informatics (ISCI), 2012 IEEE Symposium on. IEEE, 2012.

[10] Mohammad, Rami M., Fadi Thabtah, and Lee McCluskey. "Intelligent rule-based phishing websites classification." Information Security, IET 8.3 (2014): 153-160.

[11] Canali, Davide, et al. "Prophiler: a fast filter for the large-scale detection of malicious web pages." Proceedings of the 20th international conference on World wide web. ACM, 2011.

[12] Xiang, Guang, et al. "Cantina+: A feature-rich machine learning framework for detecting phishing web sites." ACM Transactions on Information and System Security (TISSEC) 14.2 (2011): 21.

[13] Canali, Davide, et al. "Prophiler: a fast filter for the large-scale detection of malicious web pages." Proceedings of the 20th international conference on World wide web. ACM, 2011.

[14] Xiang, Guang, et al. "Cantina+: A feature-rich machine learning framework for detecting phishing web sites." ACM Transactions on Information and System Security (TISSEC) 14.2 (2011): 21.

[15] WANG, Wei-Hong, et al. "A Static Malicious Javascript Detection Using SVM." strings. Vol. 40. 2013.

[16] L. Ladha and T. Deepa, "Feature selection methods and algorithms," International journal on computer science and engineering, vol 3, no 5, 2011.

[17] Hou, Yung-Tsung, et al. "Malicious web content detection by machine learning." Expert Systems with Applications 37.1 (2010): 55- 60.

[18] Cao, Ye, Weili Han, and Yueran Le. "Anti-phishing based on automated individual white-list." Proceedings of the 4th ACM workshop on Digital identity management. ACM, 2008.

[19] Huh, Jun Ho, and Hyoungshick Kim. "Phishing detection with popular search engines: Simple and effective." Foundations and Practice of Security. Springer Berlin Heidelberg, 2012. 194-207.

[20] Abela, Kevin Joshua, et al. "An automated malware detection system for android using behavior-based analysis AMDA." International Journal of Cyber-Security and Digital Forensics (IJCSDF) 2.2 (2013): 1-11.

[21] Abu-Nimeh, Saeed, et al. "A comparison of machine learning techniques for phishing detection." Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit. ACM, 2007.

OpenCV. (2015. March) Random Trees. [Online]. Available: http://docs.opencv.org/modules/ml/doc/random_trees.html

Secure Data Transmission for Health Care

Bharat Choudhary ,Rishabh Chaplot ,Sarvesh Tendulkar ,Vikas Kaul
aInformation Technology
Thakur College of Engineering and Technology
Mumbai, India
brt.chd@gmail.com,rishabh.chaplot1005@gmail.com,sarvesh.tendular08@gmail.com, kaul.vikas@thakuredu.com

Abstract-:

The Internet of things is the combination of devices, apps, sensors and network connectivity that enhances these entities to gather and exchange data Technology that enhances these features is already available but is not accessible and affordable by most of the people in developing countries such as India.

The project aims at creating an intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other hospital staff work efficiently. The system monitors patient's health standing, such as heart rate, and temperature. In case, the worth for any of those parameters exceeds predetermined essential values, the position parameters, from the hooked up WIFI module, square measure transmitted to predefined net Application web site. The data is been sensed and encrypted using the AES algorithm and transmitted to the server over the wifi module. This helps to maintain the security of the data which is been fetched and updated continuously to the web application. The project relies on open-source software, such as C Sharp (C#), ASP. Net, Visual Studio, Microsoft SQL Server. The functionality provided by these software's is enough to construct a product capable of usage for healthcare applications. For IoT hardware components such as Heartbeat Sensor, Body Temperature Sensor, Arduino, WIFI Module (ESP8266) (For Data Transmission), which can be readily available in the market. These technologies are used to build a secure data transmission for health care. The result of Secure Data Transmission for Health Care is of extreme use to patients and doctors as well. The patient's family member can check the patient's health status anytime from the comfort of their homes and visit hospitals only when they need to.

Keywords: -IOT, AES Encryption, Sensors, Security, Wifi Module

I. INTRODUCTION

Cost reduction pressures and the need for shortened in-patient stays are promoting the use of IoT Secure Data Transmission for Health Care in hospitals. Their contribution to better process management, superior flexibility and increased efficiency within hospitals is further underlining the appeal of wireless networking options for systems. Improved home care facilities and regular health updates to clinicians reduce the chances of redundant or inappropriate care. [1] It improves patient care and safety by reduction in overall costs for care. Internet of Things (IoT), gather and share information directly from patients and it also make possible to collect, record and analyze new Data Stream faster and more Accurately.

[2] As the technology for collecting, analyzing and transmitting data in the IoT continues to mature, with the help of sensors, actuators, and computing devices. [3] This provides data communication capabilities and security using the encryption algorithm which makes the sensed data safe from any type of attack. These are linked to networks for data transportation. This connected healthcare environment promotes the quick flow of information and enables easy access to diseases such as hypertension, diabetics and cardiac diseases which needs continuous monitoring. [4] This Internet of Things (IoT) is increasingly recognized by the researchers and analysts as one of the most sophisticated technologies for health monitoring and it is safety for people and it also tackled by all. [5] The ability of the devices to gather data on their own removes the limitations of human intervention and it reveals the data-automatically and send it to the doctor whenever they needed. The automation reduces the risk of error. [6]

I. OBJECTIVE AND SCOPE OF THE PROJECT

We will have an application server which will work on any platform supporting multiple departments and its sections present in-hospital major essential areas where all the emergency is required.

A. Objectives:-

1. This project aims at the planning and implementation of a coffee value; however, economic and versatile heartbeat watching and alert system exploitation IoT technology.[3]
2. It is designed in such a way that the heartbeat/pulse rate and body temperature are sensed and measured by the sensors which are being encrypted using the encryption algorithm and sends the signals to the control unit.[1]
3. Proper processing and determination of the heartbeat rate which is displayed on a Website, it then.

B Scope: -

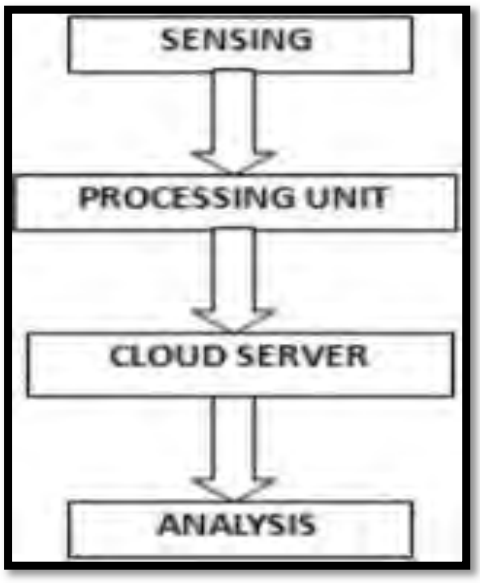
1. The Secure Data Transmission for Health Care is widely used in the healthcare system to sense the patient's heartbeat and body temperature value and to continuously update the data value on the website using the WIFI Module.[1]
2. Storing the patient's information over the cloud, which could be easily updated and accessible by the patient's family members.[6]
3. Here AES and DES encryption algorithm is been used for secure data transmission.[3]

Keywords: - Emergency, Alert, Email, Website, Threshold

II. LITERATURE SURVEY

Analysis of the current market scenario and technological capabilities is essential to the building of a project. Here we analyzed various prevalent research papers pertaining to our project as a part of the literature survey. The project
Aims at creating an intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other hospital staff work efficiently.[3]
Our literature survey was focused on the capabilities of
The choices we made for the core part of the project, namely Internet of Things (IoT) and the use of wearable technology for data collection.[2]Sensing Processing Unit 3.Cloud Server Analyses

Fig 1: - Overview



REF. NO.	PAPER TITLE	AUTHOR	YEAR OF PUBLICATION	KEY FINDINGS	RESEARCH GAPS
[1]	Patient Monitoring System Based on Internet of Things	Jorge Gómeza, Byron Oviedob, Emilio Zhumab,	2016	The system developed patient monitoring based on Internet of things, is an alternative that can be used to help patients with chronic diseases.	Health Monitoring Devices deployment is highly depended on an extensive wireless telecommunications infrastructure, which may not be available or feasible in rural areas.
[2]	Patient Monitoring Using Android Technology	Prema Sundaram	2013	Remote patient monitoring system enables the doctor who is not in the hospital to know about the details of his patient.	The lack of reliability is the most serious problem that needs to be fixed before devices and applications could be used by healthcare providers.
[3]	Patient Monitoring in Mobile Health: Opportunities and Challenges	Nilofer Mohamm ad Zadeh, Reza Safadi	2014	Applying remote medical diagnosis and monitoring system based on mobile health systems can help significantly to reduce health carcosts, correct performance.	On the other hand, it's important to take into account that not everyone owns a smartphone, and elderly people often face difficulties in using modern gadgets, such as mobile phones.
[4]	Wireless Patient Monitoring System	N. M. Z. Hashim, M. S. Sizali	2013	All the hardware design, data communication between hardware and software, communication and GUI.	Some patients are afraid that their private health data can be obtained by third parties and used for dubious purposes.

Table 1: - LiteraSurvey

III. PROBLEM DEFINITION

The Internet of Things (IoT) is a device which gathers and share information directly with transmitter and receiver, this enables the doctor to collect record and analyze new data streams faster and more accurate. The heart rate and body temperature are sensed and transmitted to the cloud server using the wifi module for monitoring purpose and it is been continuously been sensed and updated on the web application. The data which is been
is encrypted using the AES algorithm for data security

Keywords: - Transmitter, Analyze, Monitoring, Application

IV.PROPOSED METHODOLOGY

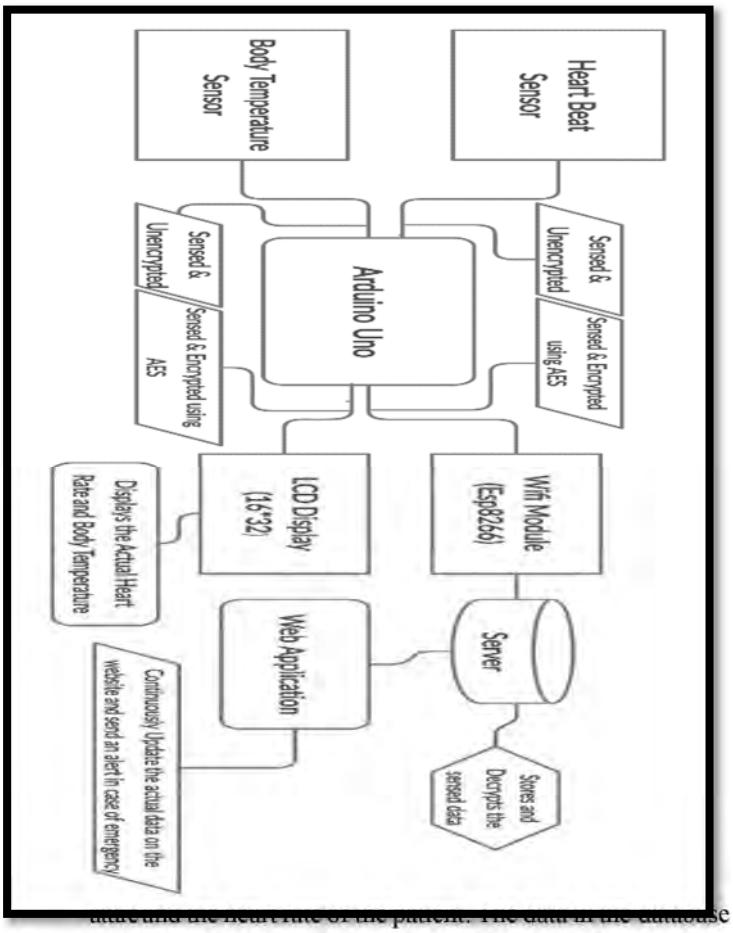


Fig 2: - Block Diagram for secure data transmission of health care

will help us to analyses whether the patient is in good condition or not and we will get to know each and every detailed information of the patient in a secured way through the secure data transmission. The website will give the notification to the patient's family members and the doctors to keep them updated regularly without getting into any tble.

A. Block Diagram

The Temperature and Pulse sensors are connected to the Arduino Temperature sensor gives the readings of room temperature and Humidity of the Patients Healthcare room through GPIO pins. Pulse Sensor gives the pulse rate readings of the heart of the patient. The code which we will implement will capture the data in encrypted format and push online to the server database.[2] There would be a web-based application where we will retrieve this sensed data and process it to show the temper

A. HEARTBEAT SENSOR

Heartbeat device is intended to provide a digital output of heartbeat once a finger is placed on that. Once the center beat detector is functioning, the beat light-emitting diode flashes in unison with every heartbeat. This digital output may be connected to the microcontroller on to live the Beats per Minute (BPM) rate. It works on the principle of sunshine modulation by blood flow through finger at every pulse.[5]

B. BODY TEMPERATURE SENSOR

The Dellas temperature generates a better output voltage than thermocouples and should not need that the output voltage is amplified; Psychological factors conjointly influence body temperature: an excited person typically has an elevated associate temperature. Traditional soma temperature, conjointly called physiological state or euthermia, is that the global temperature varies found in humans. The usual soma temperature varies usually declared as thirty-six .5–37.5 °C (97.7–99.5°F). It could be a preciseness IC temperature device with its output proportional to the temperature (in C). The electronic device equipment is sealed, and so it's not subjected to reaction and alternative processes. With this, the temperature may be measured additional accurately than with a semiconductor. It conjointly possesses low self- heating and doesn't cause quite zero.1 of temperature rise in

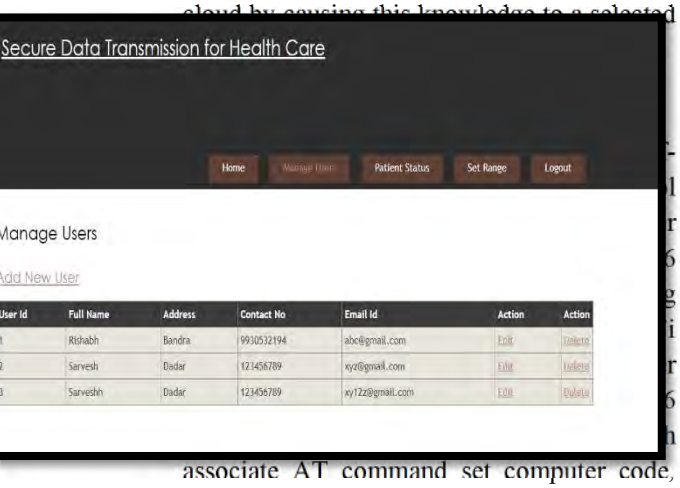
B. HEARTBEAT SENSOR

Heartbeat device is intended to provide a

D. ARDUINO

Arduino refers to associate ASCII text file physics platform or board and also the code want to program it. The Arduino UNO board unendingly reads input from these three

senses. Then it sends this knowledge to the



associate AT command set computer code, meaning, you'll merely hook this up to your Arduino device and acquire concerning the maximum amount Wi-Fi-ability as a Wi-Fi. The ESP8266 could be a cheap Wi-Fi chip with full TCP/IP stack and MCU (Micro Controller Unit) capability made by Shanghai-based Chinese manufacturer.[3]

Keywords: - MCU, ASCII, BPM, ESP8266, GPIO, Pulse, Dellas

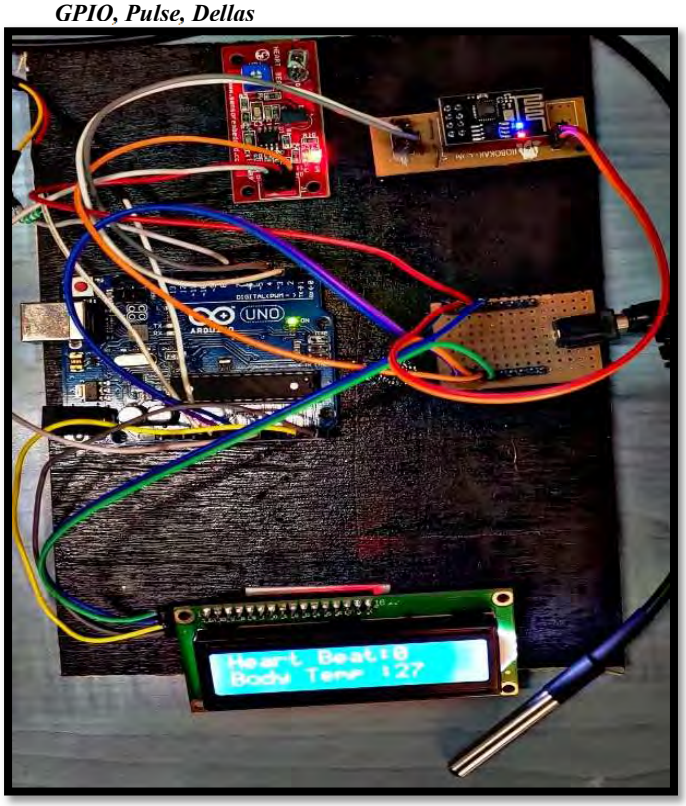


Fig 3: - Circuit Diagram

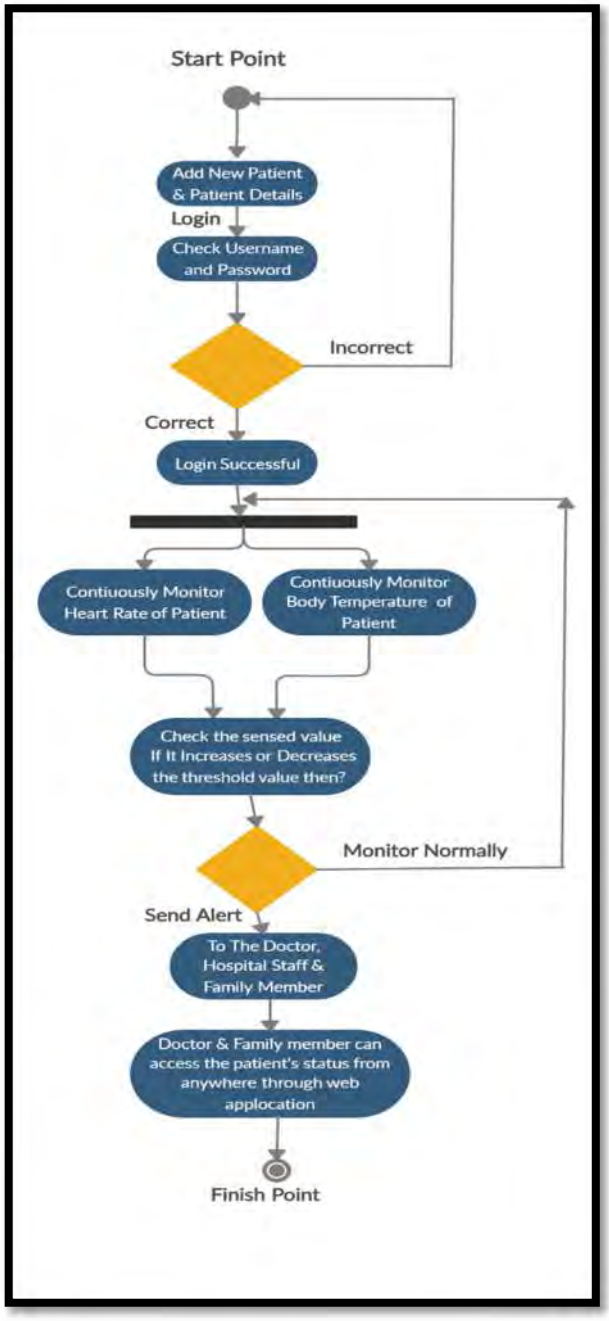
Here we have connected the temperature sensor and the pulse sensor to the

Arduino through which the detected data of

V. ACTUAL RESULTS

the patient is going to get secured through the AES encryption technique and ESP8266 is the Wi-Fi module in order to transfer the data wirelessly over the internet to the server. In case, the worth for any of those parameters exceeds predetermined essential values, the position parameters, from the hooked up WIFI module, square measure transmitted to predefined net Application web site. The data is been sensed and encrypted using the AES algorithm and transmitted to the server over the wifi module. This helps to maintain the security of the
data which is been fetched and updated continuously to the web application.[4]

F. Activity Diagram



A. Outputs

Fig 4: -- To Add New User

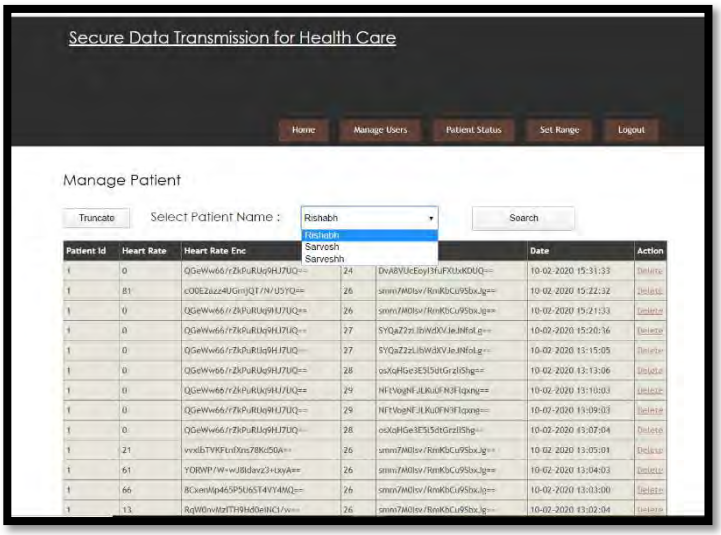


Fig 5: -- To Manage the Patients Heart Rate and Body Temp

Fig 6: -To set the threshold value according to the patient diseases.

The Pulse rate and Temperature Sensor detects the level of the humidity and temperature of patient's healthcare room and pulse rate sensor measure the level of pulses in the patient body. The result would be displayed on the webpage. This helps the concerned authority to carry on a require action.[1]

B. Outcomes

The following are the outcomes were observed as a result of implementing the product:

- Data transmitted securely through the AES encryption algorithm.
- Temperature and Pulse rate of the patient been detected.
- Transmit the information wirelessly to Webpage.
- The data can be accessed anytime and from anywhere.
- The real-time data transmission and access.
- Instant Alert to the end use.

C. Discussion of the Results

The physiological monitoring system provides different monitor units for different clinical purposes. This system architecture increases the flexibility of this system. Moreover, this system based on the Wi-Fi network is highly flexible and can be easily connected. Besides, according to the fast development of wireless technologies and the increasing communication bandwidth, we can expect that more wireless services will be provided and can be readily incorporated into the telemedicine services in the future.[5]

Keywords: - Accessed, Wirelessly, Real-Time, Flexible, Telemedicine

VI. CONCLUSIONS

This technique presents a system to upgrade existing health monitoring systems in the hospitals by providing monitoring capability and thus a better cure. This system is based upon wireless technology i.e. Cloud Computing providing low cost-effective solution. As it is wireless device, the cost of cables is reduced here. This intelligent monitoring system provides long term monitoring capability useful for the staff in the hospitals and reduces their workload. By our project we were able to measure human body temperature as well as heart rate and monitored in a remote pc or in a web application over the internet.[2] The Secure Data Transmission for Healthcare is of extreme use to patients and doctors as well. The patient can check their health status anytime from the comfort of their homes and visit hospitals only when they really need to. This can be done by using our system whose result are brought online and can be seen from anywhere around the world. Since it is a

prototype model, our system shows the almost real time values of various health parameters and emulates how the same can be implemented in the real world. The doctors can also use the log of the patient body condition to study and determine the effect of medicine or other such things.[3]

Instant availability of the patient's information is the main aim of the project. Heart beat can be sensed by using heart beat sensor which is then given to a signal conditioning circuit. This indeed is an easy, practical, inexpensive and yet very effective way for transmitting vital information to the healthcare staff and healthcare providers. The system monitors patient's health status, heart rate, and temperature. In case, the value for any of these parameters exceeds preset critical values, the position parameters, from the attached WIFI module, are transmitted to predefined Web Application Site. [6]

Keywords: - Cloud, Cost-Effective, Application Site, Secure, Transmission

VII. FUTURE SCOPE

The system can be further improved further by adding artificial intelligence system components to facilitate the doctors and the patients. The data, consisting medical history of many patients' parameters and corresponding results, can be explored using data mining, in search of consistent patterns and systematic relationships in the disease. [3] For instance, if a patient's health parameters are changing in the same pattern as those of a previous patient in the database, the consequences can also be estimated. If the similar patterns are found repeatedly, it would be easier for the doctors and medical researchers to find a remedy for the problem.[6]

RPM (Remote Patient Monitoring) device can be used

Remote patient monitoring (RPM) is a technology to enable monitoring of patients outside of conventional clinical settings (e.g. in the home), which may increase access to care and decrease healthcare delivery costs.[2]

It can be used in an RFID (Radio Frequency Identification) card.

Where the data can be stored on cloud. By just tapping the card we could retrieve the patient's past information. This system, which can be created by integrating a probable patient information system and proposed RFID application, is seen as a proper solution in cases when doctors want to have fast and automatic access to patient information and in

Particular when patients are not able to establish healthy communication.[1]

VIII. References

[1]. Robert SH Istepanian, Emil Jovanov, YT Zhang, "Guest editorial introduction to the Special Section on M-Health: Beyond Seamless Mobility & Global Wireless Health- Care Connectivity" International Journal of Information Technology in Biomedicine, Vol.8,PP.405-414,2004.

[2]. Chris Otto, Aleksandar Milenkovic, Corey Sanders, Emil Jovanov, "System architecture of a wireless body area sensor network for ubiquitous health monitoring" International Journal of Mobile Multimedia ,Vol.01, PP.307-326,2006.

[3]. Darkins, A., Ryan, P., Kobb, R., Foster, L, Edmonson, E., Wakefield, B., & Lancaster, "The systematic implementation of health informatics, home Telehealth,and disease management to support the care of Veteran patients with chronic conditions" International Journal of Telemedicine.

[4]. Chudhry, S.I., Mattera, J.A., Curtis, J.P., Spertus, J.A., Herrin, J., Lin, Z., Phillips, C.O., Hodshon, B.V., Coopers L.S., and Krumholz, H.M. "Telemonitoring in patients with heart failure", International Journal of Mass Medicals,vol.363,PP. 2301-2309,2010.

[5]. Chase, H.P., Pearson, J.A., Wightman, C., Roberts, M.D., Oderberg, A.D., N Garg, "Modem transmission of glucose values reduces the costs and need for clinic visits", International Journal of Diabetes Care,vol. 26,No.5,PP.1475- 1470,2010.

[6].Hoi Yan Tung, Kim Fung Tsang,Member, IEEE, Hoi Ching Tung, Kwok Tai Chui and Hao Ran Chi " The Design of Dual Radio Zigbee Homecare Gateway for Remote Patient Monitoring" IEEE International Conference on Consumer Electronics, Vol. 59, PP.756- 764 No. 4,201

Live location based vehicle parking system

Saurav Rathod, Stuti Shrimal, Kanishk Sonee, Shruti Mathur

Information Technology Thakur college of Engineering and Technology
Mumbai, India saurav.r99@gmail.com, stutishri@gmail.com, kanishk.sonee@gmail.com, shruti.mathur@thakureducation.org

Abstract—**In recent times the concept of smart cities has gained great popularity. Thanks to the evolution of Internet of things the idea of smart city now seems to be achievable. Consistent efforts are being made in the field of loT in order to maximize the productivity and reliability of urban infrastructure. Problems such as, traffic congestion, limited car parking facilities and road safety are being addressed by IoT. Car parking is a major issue in modern congested cities of today. There simply are too many vehicles on the road and not enough parking space. This has led to the need for efficient parking management systems Smart parking enables better and real time monitoring and managing of available parking space resulting in significant revenue generation, better urban environment and reduces fuel consumption. Present days getting a parking space in urban areas is very difficult in peak ours due to lack of parking spaces. Due to this driver stuck in traffic or looking for parking paces around the location makes traffic congestion. This causes waste of money and time. So if we have parking space information, we can plan for advance booking based on acquirement, for that we developed a prototype of cars parking management system using Internet of things. Our system proposes an loT based Smart city car Parking System on Streets which enables the user to park his vehicles in a systematic way and it reduces congestion in parking area.[3]**

Index Terms—Arduino IDE, Firebase, Node MCU

I. INTRODUCTION

nternet and its applications have become an integral part of today's human lifestyle. It has become an essential tool in every aspect. Due to the tremendous demand and necessity,
researchers went beyond connecting just computers into the
web. These researches led to the birth of a sensational gizmo, Internet of Things (IoT). Communication over the internet has grown from user-user interaction to device – device interactions these days. The Sensor Network Lab concepts were proposed years back but still, it's in the initial stage of commercial deployment. Home automation industry and Transportation industries are seeing rapid growth with Sensor Network Lab. Since most of the process is done through the internet we must have an active high-speed internet connection. The technology can be simply explained as a connection between humans' computers-things. All the equipment's we use in our day to day life can be controlled and monitored using the Sensor Network Lab. A majority of the process is done with the help of sensors in Network Lab. Sensors are deployed everywhere and these sensors convert raw physical data into digital signals and transmit them to its control center. By this way, we can monitor environmental changes remotely from any part of the world via the internet. This systems architecture would be based on the context of operations and processes in real-time scenarios. The Smart Parking Lot system works in a similar manner, with the use of a number of Ultrasonic sensors connected with a Node MCU (ESP8266) for transmission of collected data to the cloud and then to the user interface, which in our case is a website linked to cloud, using java script. The occupancy status of the parking slot is displayed on the GUI using HTML and CSS. The proposed system eliminates congestions usually faced at parking lots.[5]

II. BACKGROUND

The conventional as well as current parking systems as seen in various public places like malls, theatres, stadiums etc. have various drawbacks. The main is issue in these systems is finding or locating a parking spot that is vacant. To find a parking spot, one has to drive all around the parking lot which

can range from a few 100 sq. meters to a multi-tiered parking lot. This approach of finding a vacant parking space is extremely time consuming and inefficient. 3 Conventional parking could soon be a thing of the past. There are just too many advantages to an automatic parking system! The population is growing, and cities are running out of space. This is where the automatic parking system comes in. It doesn't only save space, but also time and money. Let's start with the disadvantages about conventional parking: The search for a parking space. Sometimes you have to drive around for ages. That's why good parking guidance system and digital signage are very helpful.

III. IMPORTANCE OF PROJECT

The various benefits and importance of attesting smart parking systems in large cities such as Mumbai are as follows:-

1. Since the entire parking system is automated, this will cut down the need of employees to manage the area, thus

cutting down the costs as well as eliminating the need of manpower.

2. Users can remotely get an overview of the parking space before actually being physically present at the area

3. Users can drive directly to the parking spot that is vacant thus eliminating congestion for the other drivers.

IV. LITERATURE SURVEY

1. "Webserver Based Smart Monitoring System Using ESP8266 Node MCU Module" BY: Dlnya Abdulahad Aziz, International Journal of Scientific and Engineering Research, ISSN 2229-5518, June 2018

The term internet of things (IOT) is considered as an important and popular technology aspect nowadays. Life style controllability based on (IOT) became considerably simpler and easier especially in the communicating approaches among the smart devices. The methodology of (IOT) allows accessing and controlling the devices anywhere and anytime. The tendency of this aspect is to communicate among various devices with respect to ESP8266 Node MCU module. The devices even can be controlled with respect to the employed ESP stations. The adeptness of data transfer among the proposed remote locations depends mainly on the behavior of the system while the security and the applicability of the system are considered more efficient. This work shares the information of air quality, Temperature and Humidity in three food stores at three remote locations with each other based on ESP module through web server.

In addition, two control actions are taken automatically in accordance with temperature and air purity output results. First, the air cooler is turned ON to cool the foods in the stores and second the air puller is turned ON to pull the contamination out of the proposed locations. Furthermore, these control actions even can be applied manually by web server administrative person. Finally, the ESP based system is classified as inexpensive project comparing to GSM module.

2. "IoT Based Smart Parking " BY: Dr. Rishi Anand, Abhirup Khanna, International Conference of Internet Of Things and Applications, IEEE 978-1-5090-0044, January 2016

In recent times the concept of smart cities have gained grate popularity. Thanks to the evolution of Internet of things the idea of smart city now seems to be achievable. Consistent efforts are being made in the field of IoT in order to maximize the productivity and reliability of urban infrastructure. Problems such as, traffic congestion, limited car parking facilities and road safety are being addressed by IoT. In this paper, we present an IoT based cloud integrated smart parking system. The proposed Smart Parking system consists of an on-site deployment of an IoT module that is used to monitor and signalize the state of availability of each single parking space. A mobile application is also provided that allows an end user to check the availability of parking space and book a parking slot accordingly. The paper also describes a high-level view of the system architecture. Towards the end, the paper discusses the working of the system in form of a use case that proves the correctness of the proposed model.

3. "Traditional Infrastructure vs Firebase Infrastructure" BY: Sonam Khedkar, Swapnil Thube, IJSRD - International Journal for Scientific Research & Development, ISSN: 2321-0613, 2017

The purpose of this study is to introduce everyone with the amazing features of Firebase that can be useful to rocket boost an Application. Traditionally to get best out of an app we need a huge, expensive, bulky methods or code to be used or implemented, long process of testing which consist of a server and large amount of complex codes, third party API's and even dedicated teams. This takes customer and developer quite far from their destiny of having an application which would be reliable, bug-free and sustainable in long run. Google's Firebase Provides such an extraordinary aid to developers/ customers which can't be possible to think of, from traditional infrastructure. This article focuses on digging into an Amazing product named Google's Firebase and its various features such as Cloud Messaging,

Authentication, Storage, Real-time Database and few more.

V. MOTIVATION

In huge metropolitan cities, the population being large enough to cause congestion at any place, parking spaces to those cities get adversely affected by the large number of vehicles that are present in that city. The main is issue in these systems is finding or locating a parking spot that is vacant. The primary motivation behind this proposed system is:-

1. The conventional as well as current parking systems as seen in various public places like malls, theatres, stadiums etc. have various drawbacks. To find a parking spot, one has to drive all around the parking lot which can range from a few

100 sq. meters to a multi-tiered parking lot. This approach of finding a vacant parking space is extremely time consuming and inefficient.

2. Conventional parking could soon be a thing of the past. There are just too many advantages to an automatic parking system! The population is growing, and cities are running out of space. This is where the automatic parking system comes in. It doesn't only save space, but also time and money

VI. OBJECTIVES AND SCOPE OF PROJECT

A. OBJECTIVES

1. User-friendly GUI to display the vacant parking spots.
2. Eliminating traffic conditions or mega-block at parking lots in some cases.
3. A smart parking system designed at a very low cost than the already existing systems.
4. Easy access to parking systems all around the city..

B. SCOPE

1. At the entry of the parking lot, a Smart Ticket can be generated to be given to the driver which would record for how much time the car was parked.
2. A secure payment gateway present at the exit of the parking lot can be introduced to help with the parking fee.
3. Moreover, a driver can get a recommendation for the nearest vacant parking space based on his or her location when the user logs in on the platform.

VII. PROBLEM DEFINITION

A. CURRENT SITUATION

In the modern society, there is an ever-increasing number of vehicles. This is leading to problems such as large urban parking lots becoming inefficient, increasing difficulty to find open spaces in busy parking lots, as well as the increasing need to devote larger areas of land for additional parking spaces. One may ask, why these problems are significant.[4]

B. PROBLEM STATEMENT

The three main problems that the increasing number of vehicles and the decreasing efficiency of modern busy parking lots are:

1. Valuable time wasted from inconvenient and inefficient parking lots on average, 3.5 - 12 minutes spent waiting for a spot in urban parking lots.

2. More fuel consumed while idling or driving around parking lots, leading to more CO2 emissions
   - Average distance travelled looking for a spot = 1.2km.
   - Average CO2 produced per car per day = 0.14 kg CO2.
   - 14kg for 1000 cars per day and 5110kg per year just for 1000 cars.
3. Potential accidents caused by abundance of moving vehicles in disorganized parking lots
   - 413 accidents occurred in public parking lots in Canada last year
   - There were 788 parked car collisions, 5 being fatal
   - 2/3 of traffic accidents in parking lots involve only 1 moving vehicle 12
     1/3 of these accidents involved 2 moving vehicles
   - Parking System Issue

Phase 1 (planning):
Perform literature survey and find all the research gaps. Phase

Phase 2 (analysis):
Prepare a report or synopsis to set a path that is to be followed after understanding the requirements and making the synopsis the development is to be started.

Phase 3 (design):

In development, the first step is to plan on how the project is to be done, the second step is to set the Ultrasonic sensors and node MCU and make the connections using wires.

Phase 4 (code , implementation):
After all the phases are completed properly, the actual development is to be started. We have created a code and is applied with the help of WiFi.

Phase 5 (testing, deployment):
The code has been rechecked and the needed connections has been made after testing and approval, apply the proposed

system.

C. FEATURES

1. Fully automated
2. User-friendly and interactive GUI.
3. Quick acquisition of vacant parking slots.
4. Minimal or no congestion caused at parking lots.
5. Cheap sensors, providing the same functionalities as a costlier sensor, thus cutting down development costs Vector Art

VIII. METHODOLOGY

A. SYSTEM ARCHITECTURE

The basic hardware components used in developing the system are Ultrasonic Sensors and NodeMCU while at the backend are present Google powered- Firebase and also Arduino IDE. When a car enters into the range of the ultrasonic sensors, the parking spot is shown vacant on the GUI.

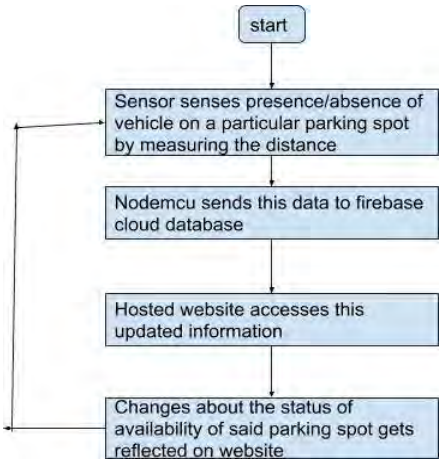


Fig 1: System Architecture (Flowchart)

B. SYSTEM PROCESS FLOW

Once the ultrasonic sensor detects the presence of a vehicle in it's vicinity, it transmits a signal to the Firebase at the backend through means of the NodeMCU (ESP8266) which is a Wi-fi module. The signal sent to the Firebase hereby communicates that the parking spot is occupied by a vehicle. This data of vacancy is displayed on the User interface which is present at the entrance of the parking lot. The data is displayed with the help of retrieving data from the Firebase using JavaScript. The user can thus have an overview of the free spaces in the parking lot.

IX. ANALYSIS AND PLANNING

A. PLANNING:

This step involves finalizing the software platforms to be used and the hardware components that will be required in the circuits.

B. DESIGNING:

In this step , the database design and the circuit connection will be finalized. Here the main architecture of the entire proposed system will also be determined.

C. BUILDING:

Here, the source code for the node MCU board as will be used to carry out basic tasks like reading data that is coming from the ultrasonic sensor and connecting to the Wi-Fi network.

D. TESTING:

After making all the connections and flashing the main source code into the node MCU , the testing phase can start where the entire system will be handling different test cases. Based on the test results

, required changes will be done in the connections or the source code. The process of testing will repeat until expected results are achieved.

E. DEPLOYMENT:

After successful test results the system can be deployed by connecting the front end site to the cloud database. The circuit will be installed at the parking spot and the users will be able to view the status of the parking lot in real time.

X. FEASIBILITY STUDY

A. TECHNICAL FEASIBILITY

This assessment focuses on the technical aspects of the project that will determine the feasibility of this proposed system. The hardware side of the project uses the extremely popular node mcu board and the ultrasonic sensor; both of which are readily available in the market. On the software side, the technologies used for cloud storage is Firebase. Firebase is Google's solution for cloud computing and it is a reliable open

source platform.

B. ECONOMIC FEASIBILITY

The proposed system's cost can vary according to the size of the parking lot. A single circuit consists of two primary components: the nodemcu board which costs around Rs.250-300 and the ultrasonic sensor which costs around Rs.50-75. Thus the final circuit will cost approximately in the range of Rs.300-375 for each parking spot, which is a cheap one time installation costs. As far as the software side is concerned, all the platforms used are free of cost.

C. OPERATIONAL FEASIBILITY

The final circuit will be placed on the ground or just below it in a small hole the size of a hockey puck and at a depth of 5-10 cms. The entire circuit will communicate with the WiFi system of the mall or the parking lot to send the data to the server. This system is fully modular as the circuits connect directly to the Router and do not depend on each other. The proposed solution is also scalable because adding a new parking spot to the system is just as easy as installing the a new circuit at that location and adding the same spot to the GUI.

XI. COMPONENTS

A. HARDWARE COMPONENTS

1 Node MCU:

Node MCU is an open source firmware developed for ESP8266 Wi-Fi chip. By exploring functionality with ESP8266 chip, Node MCU firmware comes with ESP8266 Development board/kit i.e. Node MCU Development board. Since Node MCU is open source platform, their hardware design is open for edit/modify/build. Node MCU Dev Kit/board consist of ESP8266 Wi-Fi enabled chip. The ESP8266 is a low-cost

Wi-Fi chip developed by Espressif Systems with TCP/IP protocol. [1]

2 Ultrasonic sensor:

Ultrasonic sensor is a device that can measure the distance of an object by using sound waves. It measures distance by sending out a sound wave at a specific frequency and waits for that sound wave to bounce back. By recording the time taken between the sound wave being generated and the sound wave bouncing back, it is possible to calculate the distance between the sensor and the object.

B. SOFTWARE PLATFORMS

1 Firebase:

Firebase provides a real-time database and backend as a service. The service provides application developers an API that allows application data to be synchronized across clients and stored on Firebase's cloud. The company provides client libraries that enable integration with Android, iOS, JavaScript,Java, Objective-C, Swift and Node.js applications. The database is also accessible through a REST API and bindings for several JavaScript frameworks such as AngularJS, React, Ember.js and Backbone.js. The REST API uses the Server-Sent Events protocol, which is an API for creating HTTP connections for receiving push notifications from a server. [3]

2 Arduino IDE:

The open-source Arduino Software (IDE) makes it easy to write code and upload it to the board. It runs on Windows, Mac OS X, and Linux. The environment is written in Java and based on Processing and other open-source software. This software can be used with any Arduino or Node MCU board.

XII. EXPECTED RESULTS

A. OUTPUTS

Successfully implemented the Smart Parking System in an economical way. The ultrasonic sensor measures the distance between the sensor itself and the vehicle parked above it. The vacancy data is sent to the Firebase through the NodeMCU and then onto the web application for the user to view.

Phase 1: Analysis

- Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase.

Phase 2: Planning

- Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done

Phase 3: Designing

- Proposing design architecture of Smart Parking System for implementing the surveyed case studies and synopsis.

B. *OUTCOMES*

The following are the outcomes were observed as a result of implementing the product:

- Transmission of data wirelessly from the NodeMCU to the Firebase at the back end.
- Status of the real-time occupancy is presented to the users on the web application which can be accessed even when the user is not present at the parking premises

C. *FUTURE SCOPE*

The main objective of the proposed system has been fulfilled, yet, there are many more things that could be integrated with the already existing system to make it more responsive and smart. The following features could be integrated in the future:-

- An App based application could be designed instead of the existing website
- Geo-location of the parking lots can be provided to the users to help them find the nearest parking lots instead of parking on the streets. This can be implemented using the Google maps API at an initial level.
- A secure payment gateway can be developed at the exit on basis of the period of time the user had parked his/her vehicle in the premises. The time can be recorded by providing a randomly generated key with the timestamp printed on a parking slip provided at the entrance which has to be scanned at the exit for commencing payment.[6]

XIII. CONCLUSION

A In this project we can herby conclude that we can make the best out of ultrasonic sensor and node mcu , that is using it for monitoring tight parking spaces in a busy metropolitan city like Mumbai and it can be used all throughout the world for better space allocation and avoiding stockpiling and traffic of cars. This mainly focuses on the best allocation and indicates a user of this service whether the space is occupied or not and does inform on a prior basis about the required space. It will be able to re direct the user based on the availability of the parking space. The basic Idea is that the sensor detects if there is a vehicle parked on a given area and the sends the information across using Wifi technology and informs the system about the space. This helps to avoid nuisance and keeps the disturbances at lows. Thus this technology in tern helps the users to get through their busy day at a faster pace by allocation of spaces for their vehicles at any given instances. The technology is much cheaper and easier to use than the existing systems that exit today. They are much more expensive and need a learning curve for someone to understand the working and people are required for maintenance. This solves all the problems of a parking allocation and provides the ease of use to every user, efficiently.

References

[1] Webserver Based Smart Monitoring System Using ESP8266 Node MCU Module (2018), Dlnya Abdulahad Aziz, International Journal of Scientific and Engineering Research, Issn 2229-5518.

[2] IoT based Smart Parking System (2016), Dr. Rishi Anand, Abhirup Khanna, International Conference of Internet Of Things and Applications, IEEE 978-1-5090-0044 [3]Traditional Infrastructure vs Firebase Infrastructure (2017), SonamKhedkar, SwapnilThube, IJSRD - International Journal for Scientific Research & Development, ISSN: 2321-0613. [4]https://www.smartparking.com [5]https://www.forbes.com/sites/pikeresearch/2017/01/26/sma rt-parking/

[3] https://iot.electronicsforu.com/expert-opinion/smart-city-solutions-smart-parking-lots/

Anti-Phishing Mailbox

Vivek Singh, Vinay Singh, Shivam Vishwakarma, Rahul Neve
Thakur College of Engineering and technology, Mumbai, India
viveksingh3421@gmail.com, vinaysingh1093@gmail.com, shivamvishwakarma84@gmail.com, rahul.neve@gmail.com

Abstract- **Phishing has created a serious threat towards internet security. Phish e-mails are used chiefly to deceive confidential information of individual and organizations. Phishing e-mails entice naïve users and organizations to reveal confidential information such as, personal details, passwords, account numbers, credit card pins, etc. Phisher spread spoofed e-mails as coming from legitimate sources, phishers gain access to such sensitive information that eventually results in identity and financial losses. In this research paper, exhaustive study is done on anti-phishing mechanism from year 2002 to 2014. A comparative analysis report of anti-phishing detection, prevention and protection mechanisms from last decade is listed. This comparative analysis reports the anti-phishing mechanism run on server side or client side and which vulnerable area is covered by it. The vulnerable area is divided into three categories on the basis of email structure.**

The number of vulnerabilities covered by existing anti-phishing mechanisms are listed to identify the focus or unfocused vulnerability. . This research paper could be said as tutorial of an existing anti-phishing research work from decade. The current work examines the effectiveness of the tools and techniques against email phishing. It aims to determine pitfalls and vulnerability of anti-phishing tools and techniques against email phishing.

Keywords- Anti-phishing, Mail-Security, Phishing.

I. INTRODUCTION

Phishing is defined as the fraudulent acquisition of confidential data by the intended recipients and the misuse of such data. The phishing attack is often done by email. An example of Phishing; as if e-mail appear to be from known web sites, from a user's bank, credit card company, e-mail, or Internet service provider. Generally, personal information such as credit card number or password is asked to update accounts. These emails contain a URL link that directs users to another website. This site is actually a fake or modified website. When users go to this site, they are asked to enter personal information to be forwarded to the phishing attacker [1, 2].

Phishing is often used to learn someone's password or credit card information. With the help of e-mail prepared as if coming from a bank or official institution, computer users are directed to fake sites. In general, the information that is stolen by a phishing attack is as follows:

- User account number
- User passwords and user name
- Credit card information
- Internet banking

The Anti-Phishing Simulator, which is designed to prevent serious threats like this, catches malicious e-mails arriving at e-mail addresses integrated into the system. This system also provides URL based control. The system evaluates the keywords included in the existing database and thus determines the contents of the mail [3, 4].

Phishing Attack- Phishing sends a fraudulent transmitter that appears to come from a real source. It is usually done by e-mail. The aim is to steal sensitive data such as credit card and login information or to install malicious software on the victim's machine. Phishing is a common type of cyber-attack that everyone must learn to protect themselves. Phishing is start with a fake e-mail or other type of transmission designed to attract a victim. In this type of attack, the message appears to come from a trusted source. If the attacker is deceiving the victim, it is mostly encouraged to provide confidential information in a fraud web site. Sometimes malware is downloaded to the target computer. Attackers provide financial gain by having their victim's credit card information or other personal data. Sometimes, phishing e-mails are sent to retrieve login details or other details of employees to use for an advanced attack against a specific company. Cyber-attacks, such as Advanced Permanent Threats (APT) and ransom software, usually start with phishing.

In a phishing attack, attackers can use social engineering and other public information resources, including social networks like LinkedIn, Facebook and Twitter, to gather background information about the victim's personal and work history, interests and activities. With this pre-discovery, attackers can identify potential victims' names, job titles and email addresses, information about the names of key employees in their colleagues and organizations. This information can then be used to prepare a reliable e-mail. These attacks, including attacks by advanced persistent threat groups, usually start with an e-mail containing a malicious link or attachment. In this type of attack, the most popular vulnerability or clickable phishing environments have been identified as the most popular Facebook feeds. When phishing attacks are created, they are often used for unrealistic news, such as those created around major events, holidays and anniversaries. Usually a victim receives a message that appears to have been sent by a known person or organization. The attack is carried out via a malicious file injection that includes phishing software or through links to malicious websites. In either case, the goal is to direct the user to a malicious website to install malicious software on the device or to trick them into disclosing personal and financial information, such as victims, passwords, account IDs, or credit card

details. A successful phishing message is usually shown from a well-known company; it is difficult to tell from the original messages: in 978-1-5386-3449-3/18/$31.00 ©2018 IEEE phishing e-mails, company logos and other descriptive graphics and data collected from the company. As with other link manipulation techniques, the use of subdomains and misspelled URLs (often spelling mistakes) is common. Phishing attackers use JavaScript to place a legitimate URL of the URL onto the browser's address bar. The URL generated by navigating through an embedded link can also be modified using JavaScript. Defense against phishing attacks should begin with training and informing users to identify phishing messages; but there are other strategies that can reduce successful attacks. For example; a network gateway e-mail filter can capture many targeted phishing e-mails and reduce the number of phishing e-mails reaching users' inboxes [3]. Figure 1 illustrates the process of performing phishing attacks.

II. BACKGROUND

Late years have seen, phished messages as a top-notch device for trickiness of online clients into uncovering delicate data. Right off the bat phishers assemble data by 'social building' (making laws or utilizing different techniques to impact popular feeling and take care of social issues or improve social conditions). They at that point make phished messages and devour them through various specialist organizations like Gmail, Yahoo, Rediff mail, hot mail, and so on. On a very basic level, a phished email is a copy of a general, real email from a confided in source (individual or association). A phished email can be, (a) Textual, offering unique plans at worthwhile rates, fortunate victor rewards, and so forth to draw individual data; (b) Embedded with a url diverting the online client to counterfeit pages, requesting that the beneficiaries do likewise as in point (a); (c) Rooted with a pernicious connections, i.e., tainted PDFs or MS Office reports, et al that, after being downloaded, increase undignified access to private data

III. IMPORTANCE OF THE PROJECT

Phishing sends a fraudulent transmitter that appears to come from a real source. It is usually done by e-mail. The aim is to steal sensitive data such as credit card and login information or to install malicious software on the victim's machine. Phishing is a common type of cyber-attack that everyone must learn to protect themselves. Phishing is start with a fake e-mail or other type of transmission designed to attract a victim. In this type of attack, the message appears to come from a trusted source. If the attacker is deceiving the victim t is mostly encouraged to provide confidential information ii a fraud web site. Sometimes malware is downloaded to the target computer. Attackers provide financial gain by having their victim's credit card information or other personal data. Sometimes, phishing e-mails are sent to retrieve login details or other details of employees to use for an advanced attack against a specific company. Cyber-attacks such as Advanced Permanent Threats (APT) and ransom software, usually start with phishing

IV. OBJECTIVES AND SCOPE OF THE PROJECT

This project aims throwing light on the various dimensions of email phishing. Through this exhaustive study and analysis, it is identified that the researchers were focused mostly on page content vulnerabilities which was mostly used to trick naïve users. Domain vulnerabilities and code scripting vulnerabilities are prone to bigger scams and uses advanced technology to trick naïve as well as proficient users. It has been noticed that lot of work done with regard to reducing phishing, considerable work is being done to provide a safe and secure email and internet service, the same can also be said to be true for the other the fence but still is required to stop from root. This research paper gives a background and a deep literature study on anti-phishing Shweta Shank war et al, International Journal of Advanced Research in Computer Researchers are here recommended to patch all the dimensions of phishing. However, with consistently improving technology and the research that goes into it, the phishing detection and removal techniques might lead to a future where phishing has ceased to exist.

V. SUMMARY

Anti-Phishing mailbox aims to control the security of information and to prevent Infringements, to check whether spam is available from the current database, to enable the use to create his own spam list, and to check whether the incoming mail has dangerous content. The inclusion of the mail account to be protected in the system with this module, the user will also control and communicate without having to open the mail. Figure shows a screenshot of the spam mails detected in the mail system. With this module, it is possible to determine the classification results of keywords and passages in the

database by Link Guard algorithm. The panel where the URLs in the mail are checked and the fake URL is detected.

VI. DESIGN AND IMPLEMENTATION:

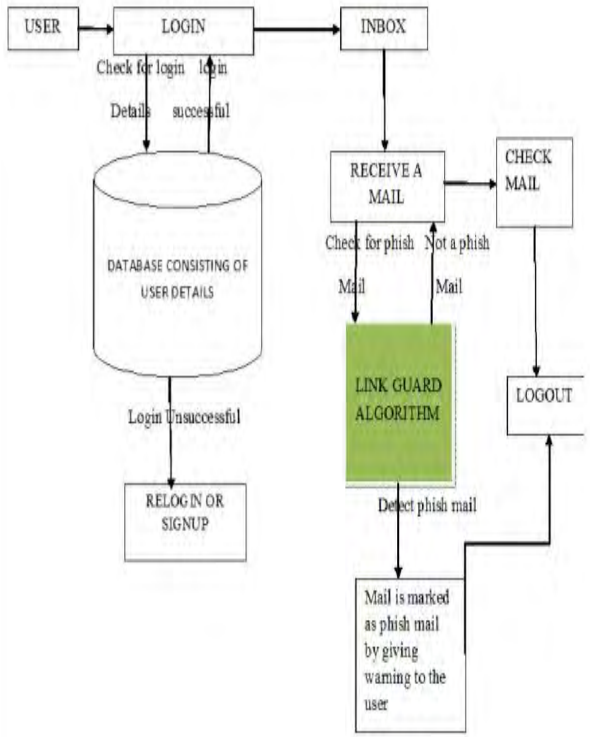
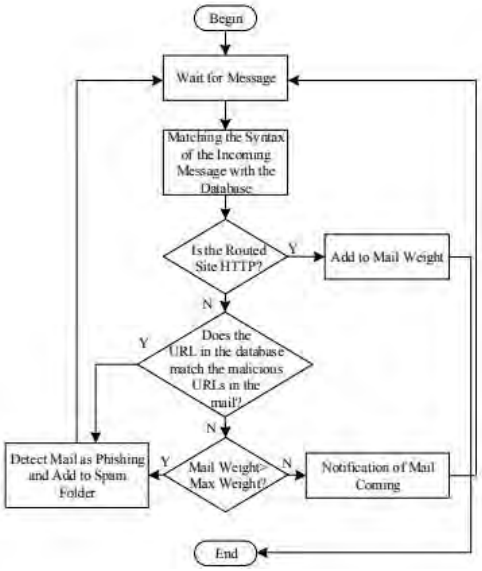


Figure: Block diagram

Flowchart

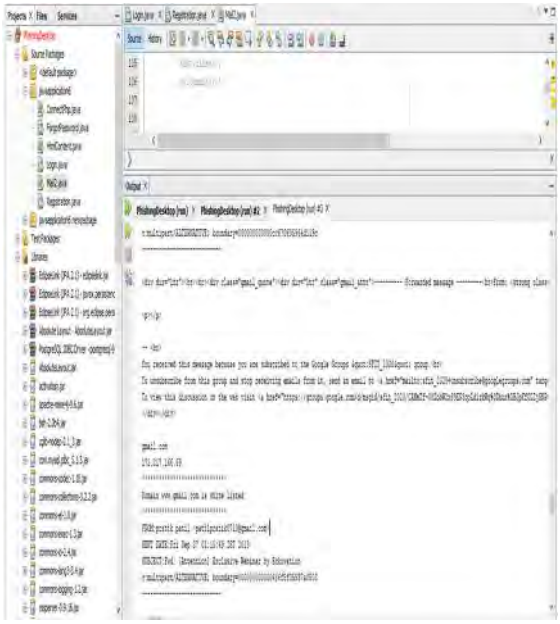


GUI:

Figure: Login Page

Figure: Registration

VII. RESULT:



Software – Net Beans IDE is a free, open source, integrated development environment (IDE) that enables you to develop desktop, mobile and web applications. The IDE supports application development in various languages, including Java, HTML5, PHP and C++. The IDE provides integrated support for the complete development cycle, from project creation through debugging, profiling and deployment. The IDE runs on Windows, Linux, Mac OS X, and other UNIX-based systems. The IDE provides comprehensive support for JDK 7 technologies and the most recent Java enhancements. It is the first IDE that provides support for JDK 7, Java EE 7, and Java 2. The IDE fully supports Java EE using the latest standards for Java, XML, Web services, and SQL and fully supports the Glassfish Server, the reference implementation of Java EE.

PHP started out as a small open source project that evolved as more and more people found out how useful it was. Rasmus Lerdorf unleashed the first version of PHP way back in 1994.

- PHP is a recursive acronym for "PHP: Hypertext Preprocessor".
- PHP is a server side scripting language that is embedded in HTML. It is used to manage dynamic content, databases, session tracking, even build entire e-commerce sites.
- It is integrated with a number of popular databases, including MySQL, PostgreSQL, Oracle, Sybase, Informix, and Microsoft SQL Server.
- PHP is pleasingly zippy in its execution, especially when compiled as an Apache module on the Unix side. The MySQL server, once started, executes even very complex queries with huge result sets in record-setting time.
- PHP supports a large number of major protocols such as POP3, IMAP, and LDAP. PHP4 added support for Java and distributed object architectures (COM and CORBA), making n-tier development a possibility for the first time.
- PHP is forgiving: PHP language tries to be as forgiving as possible.
- PHP Syntax is C-Like.

VIII. Outcomes

The Anti phishing mailbox detects all the phishing mails and blacklist all and allow all type mails which is white listed because of this Mailbox the user can only access the mails which is necessary and feasible. Email is one of the most significant specialized strategies. Expanded spam messages cause traffic blockage, diminished profitability, phishing and this is a significant issue as far as the universe of data. The quantity of spam messages is expanding each year. Consequently, spam email separating is a significant, important and testing issue. Because of the quick spread of phishing assaults, various methods for assurance have been created. Genuine and counterfeit website pages are some of the time very hard to tell from the way that phony pages are the equivalent as far as structure. The steady development of email clients has brought about undesirable messages getting so boundless. Existing server and customer side hostile to spam channels are

being utilized to recognize various highlights of spam messages. Be that as it may, some viable stunts have been created with the expansion of spam senders' spam content as advanced pictures, pdf and word; this augmentation has rendered it insufficient for current systems dependent on examining computerized message in the body zones of the email. A large portion of the work methodology proposed in the examination gives an against spam separating approach dependent on information mining procedures that characterizes spam and phishing messages. The viability of these methodologies is assessed on the expansive body of the basic content informational collection and the content implanted picture informational collection.

"Hostile to Phishing Simulator" gathers phishing and spam messages at a typical point. Messages in the spam box, it enables you to control the "spam box" at whatever point you need. The individuals who are in fact qualified by the "URL Control" highlight will have the option to look at the connection address via the post office in more detail. Later on, it is planned to dissect mail content all the more completely with fundamental content mining by expanding the spam watchword database considerably more. It is additionally meant to get increasingly exact outcomes and characterization with fake neural systems.

IX. Conclusion

Phishing is the try to accumulate touchy facts such as usernames, passwords, and credit score card information (and occasionally, circuitously, cash), often for malicious motives, through masquerading as a honest entity in an digital communication. Now days it has emerge as very serious. There are many techniques to solve those problems. But humans may also don't aware of the seriousness of phishing. Periodical updating of anti-phishing gear or software's in their personal structures may useful to relaxed their personal data and credentials. This examine may additionally give the attention about the phishing issues and answers. Anti-Phishing mailbox aims to control the security of information and to prevent Infringements, to check whether spam is Available from the current database to create his own spam list, and to check whether the incoming mail has dangerous content. The inclusion of the mail account to be protected in the system with this module, the user will also control and communicate without having to open the mail.

References

[1] International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Dr. Radha Damodaram "STUDY ON PHISHING ATTACKS AND ANTIPHISHING TOOLS" 2016

[2] Jyoti Chhikara , R.V Dahiya , Neha Garg , Monika Rani "Phishing & Anti-Phishing Techniques" 2013 International Journal of Advanced Research in Computer Science and Software Engineering

[3] Muhammet Baykara , Zahit Ziya Gürel "Detection of phishing attacks"2018 national Conference on Electrical, Computer and Communication Technologies (CECCT, Coimbatore, p 1-3 2015.

[4] A K Sharma and R. Yadav, "Spam Mails Filtering Using Different Classifiers with Feature Selection and Reduction Technique" 2015 Fifth International Conference on Communication Systems and Network Technologies, Grvaliorpp 1039-1093. 2015

[5] T. Vyas, P. Prajapati and S. Gadhva Study on phishing attacks and anti-phishing tools l. "A survey and evaluation of supervised machine learning techniques for spam e-mail filtering. 2015 national Conference on Electrical, Computer and Communication Technologies (CECCT, Coimbatore, p 1-3 2015

Smart Blood Storage Monitoring and Management System using IoT

Tanisha Gupta, Janhavi Patil, Sidddhant Manjrekar, AbdulRehman Shaikh
Information Technology Thakur college of Engineering & Technology Mumbai ,India
tanishag78@gmail.com, janhavip61198@gmail.com, siddhantmanjrekar042@gmail.com, abdulshaikh2504@gmail.com

Abstract - **The unit which administers and manages the requisition and distribution of the blood is named as a blood bank. The main objectives of the blood banks are providing blood to the patients with minimal blood transfusion error. The blood is very important medical supplies so it should be managed well. As the blood bank management consists of a number of manual steps, therefore it will become difficult for the blood banks to provide a high level of accuracy, reliability, automation in blood storage and transfusion process. The system proposed is divided into three segments, the first segment consists Temperature sensor, IR sensor nodes which is installed in rack of blood bank, and the GSM Module for sending request of blood to the donors and blood banks all these are interfaced with Arduino Mega. Second segment consists of wi-fi module for data transfer to the server and third segment is displaying the status of available blood stock. All the real time status relates to the available blood stock of the blood bank is are displayed on web page, so that the blood seeker can get the blood from their nearest blood bank.**

I. INTRODUCTION

The process of managing the blood bag that is received from the blood donation events needs a proper and systematic management that is done by the blood banks. As the blood is related to someone's life so the bags of must be handled with care and treated thoroughly. The blood consists of different blood constituents and every constituent of blood is having a specific function. Statistics say that in our India every two seconds someone needs a blood transfusion. The Blood transfusions are used for trauma victims - due to accidents and burns - heart surgery, organ transplants, women with complications during childbirth, newborns and premature babies, and patients receiving treatment for leukaemia, cancer or other diseases, such as sickle cell disease and thalassemia [6].The different blood constituent is applied to the patient for different type's disease given in table1.

India today faces a blood shortage of 10% relative to its blood requirements. It means that we require covering a shortfall of over 12 lakhs of blood units. Given that the eligible donor population of our India is more than 512 million, this deficit is surprising. Every day more than 1200 road crashes occurring in our India, 60 million trauma induced surgeries are performed in the country every year. Near about 230 million major operations, 331 million cancer-related procedures like chemotherapy and 10 million pregnancy complications all these require the blood transfusion [9]. Besides all this, the patients which are for sickle cell anaemia, thalassemia and haemophilia require large quantities of blood daily. Another reason for the shortage of the blood supply is the ban on payment to blood donors that was enacted in 1995. Previously near about onethird of the blood supply was from paid donors. But with the ban, both private and government hospitals have faced a larger shortage in the blood supply.

Table -1: Different Blood Constituents with their uses for different diseases and there functions

Blood Constituents	**Functions**	**Diseases**
Plasma	Medium in which the blood cells are transported around the body	Burn patients, Shock, Bleeding disorders
Red blood cells	Carries oxygen	Surgery, Any blood loss , Blood disorders such as sickle cell
White blood cells	Part of the immune System	Infectious disease and foreign invaders
Platelets	To facilitate blood clotting	Cancer treatments, Organ transplants ,lower platelet counts, suffering from leukemia

The current system that is using by the blood bank is manual system. With the manual system, there are problems in managing the records related with blood stock. There is no centralized database of volunteer donors. So, it becomes really tedious for a person to search blood in case of emergency. Without an automated management system, there are also problems in keeping track of the actual amount of each and every blood type in the blood bank. In addition, there is also no alert available when the blood quantity is below its par level or it become zero. Healthcare as a sector is a definite beneficiary of the increasing adoption of technologies such as the Internet of Things. One of the areas where IoT can be used is blood bank management. In our country thousands of litters of blood get wasted every day due to less than adequate management practices, thousands of litres of this precious resource get wasted every day across the country. Such a wastage is occurring during a day and age when healthcare providers are battling an acute shortage of blood units for various medical

procedures. Given this backdrop, the significance of deploying technologies that encourage efficient management of blood banks cannot be understated. The purpose of the blood bank management system is to simplify and automate the process of searching for blood in case of emergency and maintain the records of blood stocks in the bank.

II. METHODOLOGY

The proposed system is an embedded system which will closely monitor the available status of the blood and temperature of the refrigerator which is present inside the blood bank. The proposed system includes service domain and administration domain. Service domain has blood bank that has sensing unit installed to it which has array IR sensor, temperature sensor, Arduino mega as a gateway with Wi-Fi module using wireless protocol for wireless communication between blood bank and cloud, RTC for real-time operation.

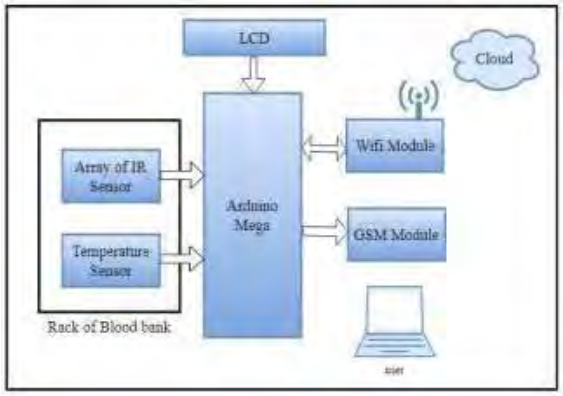


Figure -2: Block diagram of IoT Based blood bank system

In Administration domain information transferred from the blood bank is analyzed and processed. It includes webpage hosted on a server, for analysis and processing of the received data from the blood bank, depending on which corresponding notifications are displayed. It uses HTTP protocol for secure communication between server and blood bank in an application layer of IoT. For the more efficient operation, the called blood bank locations are located using maps. The two domain of proposed system is given below.

2.1 Service domain:

The service domain consists the hardware part of the system. It includes IR sensors, Temperature sensor, Arduino Mega, LCD, GSM module and ESP Module. The IR sensors are used for providing information of the available bloodstock, if the blood bottle is placed at rack then the LED rays of IR sensor gets reflect back then signal from IR to Arduino has been send. If the bottle is not placed in such condition the Rays not gets reflect back and its show the bottle is not present. For displaying the temperature of the refrigerator the temperature sensor is used which is display on the LCD. If the bloodstock is not available in such condition the Request message for donating blood is sent to the blood donor as well as one request message is sent to the nearest blood blank to provide the blood if they have availability of specific blood group of blood. All the data related to available bloodstock is displayed on the website using ESP module.

2.2 Administration domain:

In this domain, the Blood bank information, Blood status information, location of blood bank, the data received is collected. To achieve this web hosting is used where server is used. Server manages database for the information received from service domain i.e. values for each sensor for blood unit of each blood bank.

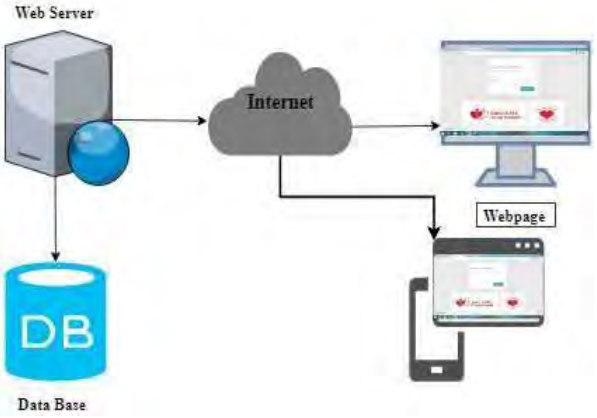


Figure -2.2: Administrator Domain for Smart Blood Bank System

It consists the IoT part, as ESP module sends the information of available blood is sent to the cloud. The cloud is connected to the website so the website will show all available bloodstock on 24*7 basis. So initially user has to select their region from options of the website the according to that the website will display all the blood bank of that region. And that blood bank provides real-time available bloodstock on the webpage. Also, the direction of that specific blood bank is provided on the google map for that one tab is provide as locate on the map.

III. RESULTS

Blood bank centers and various organizations needs to register on website of blood bank. So that each blood bank is getting connected with the cloud so that data of available blood stock of every data registered blood bank is displayed on the webpage. Also the log related with blood stock also available on the webpage so that it become easy to cross check the data of the blood stock.

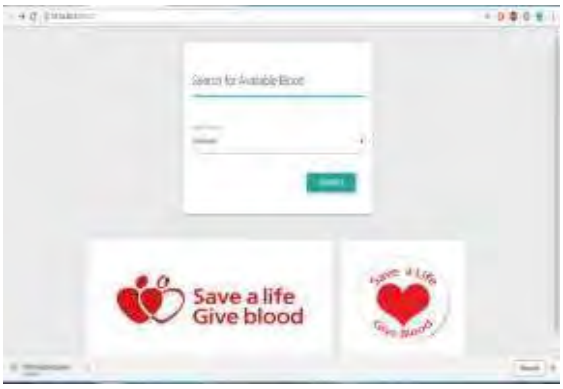


Figure - 3.1: Homepage of website

The results of the system going to be displayed on the webpage hosted over server i.e. the real time data frames sent in bytes by service domain module with the HTTP protocol to server. By using the domain name or IP address of the webpage in browser the real time data is seen on the computer or mobile that is connected to internet. Initially from homepage blood seeker has to choose the specific district of Maharashtra state, then after choosing the specific district of Maharashtra the page will show the status of available blood stock in each blood bank of specific region.

The webpage display real time information of available blood stock over. This saves time and is useful for minimizing the route in turn saving the time and fuel. The Screen shot of websites home page is shown above in figure 3.1.

After choosing the district the webpage directed to the next webpage, which is displaying available blood stock in the all blood banks. The Figure. 3.2 Shows the search result of Amravati District, it is showing four blood bank status located at the Amravati region. It is displaying status of four blood group as A, B, AB and O. Also the route of the Blood bank also provided on the webpage.

Figure -3.2: Search Result of Website

After finding availability of the blood the users need location of blood bank. So the webpage providing google map facility so that the user will finding the exact location of blood bank and also user will gets the exact distance of that blood bank so it will become easy to choose blood bank which is placed at the minimum distance from them. The Figure. 3.3 Shows Map it locate the blood bank, so any blood seeker can get the route of the blood bank. The direction toward any nearest blood can be obtained from the Google map.

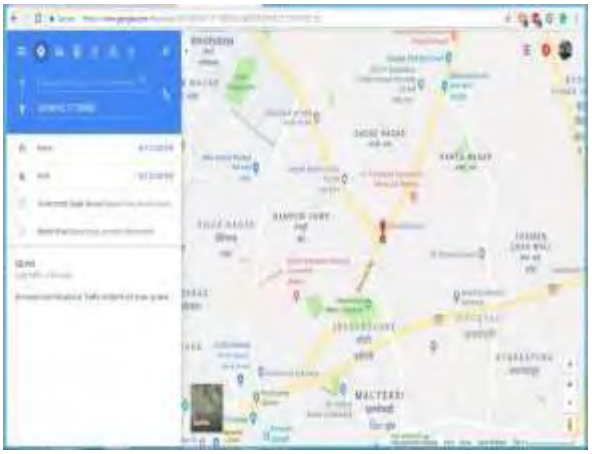
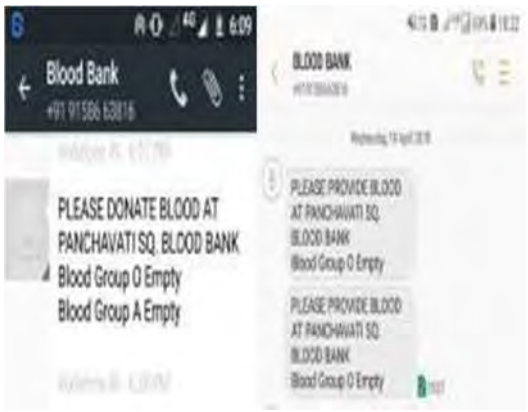


Figure -3.3: Location provided on Google Map

The System sends a short message to staff of other blood bank and blood donor's mobile phone when system senses the blood stock is below than the acceptable limits. When any blood stock reaches to zero in that case two different requests are send form the respective blood bank to full filling the blood stock. One request is send to the nearest blood bank and other request is send to the blood donor shown in figure 3.4



(a) **(b)**

Figure -3.4: (a) Screen shot of blood request Received by Blood Donor
(b) Screen shot of Blood Request Received by Nearest Blood Bank

Also the log file of the available blood stock is also obtained from the website of blood bank so that is will become easy for the blood bank employee to cross the data of the blood stock. And they can easily download that log file from the webpage itself shown in figure 3.5. It shows data of all the four blood banks.

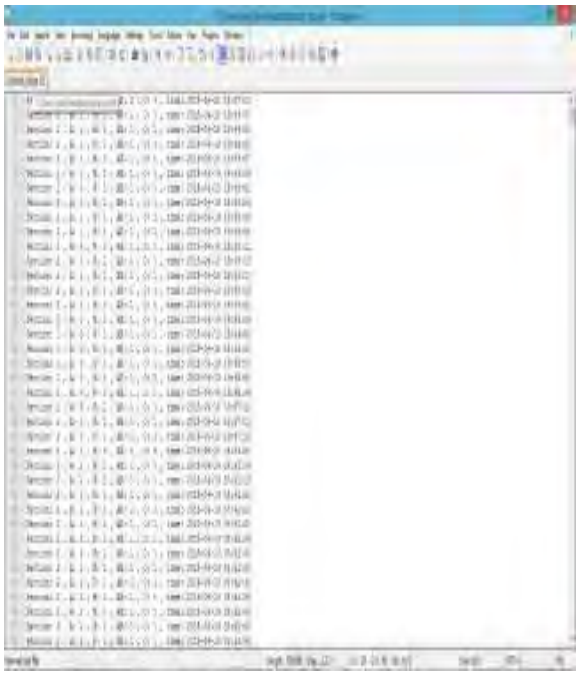


Figure -3.5: Log file of available blood stock

IV. CONCLUSION

Growing population has increased the need for the blood supply for various diseases. In every two seconds, some person required blood transfusion and currently India facing problem of the blood shortage. To address the problem an effective system is designed using the Internet of things. The system provides a methodology to fulfil the requirement of blood to the patients/victims without rushing to the blood bank to know the availability of the blood. An IR Sensors are connected to the Arduino board which continuously monitors the status of the available bloodstock. The output data provided by the Arduino is displayed on the webpage using the wifi module so anyone accesses the website and obtained the information of available bloodstock in real time. It will reduce the manpower required at the blood bank to update the online data also reduces the efforts of blood seeker of searching bloodstock at each blood bank. When bloodstock reaches to zero system helps to send a request message to the donor and nearest blood bank. By using IoT the real-time available bloodstock is displaying on the website it minimizes the efforts of blood seeker.

V. FUTURE IMPLEMENTATION

There are three different topics related to the Internet of Things concept that has major future scope in terms of improvement and research: efficiency, scalability and quality of service. The developed system is prototype version which provides the information about the available bloodstock. It consists of an array of IR sensors that covers the minimal area. To cover large blood packet the IR sensors can be replaced by the load cell. It will cover the large blood packet and help efficiently to provide information on bloodstock on a real-time basis. Also, one alert system using RFID Tag with smart temperature sensor can be also added to the system for indicating the expired blood packet and its temperature.

In the future, this work can be extended in the context of total blood banks of a country. Big Data analysis can be done on the gathered data from Region. So that the specific domain names can also be allocated to the blood bank site. Also on the website of blood bank, the login id and passwords can be provided to the blood seekers so that they can be made the online request for the blood units.

References

[1]. J. S. Lee, Y. W. Su, and C. C. Shen, "A comparative study of wireless protocols:Bluetooth, UWB, ZigBee and Wi-Fi," in Proc. IEEE 33rd Annu. Conf.IECON, pp.46-51, Nov. 2007

[2]. Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari,and Moussa Ayyash, "Internet of Things: A Survey on Enabling Technologies,Protocols, and Applications" IEEE Communication Surveys Tutorials, Vol.17, No. 4, Fourth Quarter 2015,pp. 2347-2377.

[3]. G. Muddu Krishna; S. Nagaraju(2016), "Design and implementation of short message service (SMS) based blood bank", 2016 International Conference on Inventive Computation Technologies (ICICT)

[4]. Do-Sung Kim, Sun K. Yoo, HO Kim, BC Chang ,HS Bae , SJ Kim(2007), "Location Based Blood Bag Management using active RFID and Ubiquitous Sensor Network" ,6th

[5]. International Special Topic Conference on ITAB, 2007, Tokyo

[6]. Wijai Boonyanusith; Phongchai Jittamai (2010), "The Development of Web-Based System for Blood Requisition within Blood Supply Chain", 2010 Seventh International Conference on Information Technology: New Generations

[7]. Muhammad Arif; S. Sreevas; K. Nafseer; R. Rahul(2012) "Automated online Blood bank database", 2012 Annual IEEE India Conference (INDICON)

[8]. Adarsh N, Arpitha J, Md. Danish Ali, Mahesh Charan N, Pramodini G Mahendrakar (2014), "Effective Blood Bank Management Based On RFID in Real Time Systems",

[9]. International Conference on Embedded Systems - (ICES 2014)

[10]. Anish Hamlin M R, Albert Mayan J (2016), "Blood Donation And Life Saver-Blood Donation App", 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)

[11]. https://www.ndtv.com/health/why-is-there-a-bloodshortage-in-india-1712012

IOT Based Green House Monitoring and Control System

(TCET-IT, Student) Mumbai, India.
yashsinghania15@gmail.com, ayushsharma0009@gmail.com, muskantiwari0697@gmail.com, purvisankhe@gmail.com

Abstract— **In the ongoing time, with the fast flood of development in territories of computerization and digitalization, our lives are improving as nearly everything is experiencing motorization which is supplanting the crude manual, escalated work frameworks. Today, as Internet has become an indispensable vital piece of our everyday lives, all the electronic gadgets likewise should be expedited the system. This is the primary point of Internet of Things (IoT), which is one of the new up and coming advancements that aides in setting up association, control, and the executives of independent insightful articles which are associated with Internet by means of various capacities and conventions.**

In our task the fundamental goal is to concentrate on consolidating IoT innovation in the field of farming, explicitly in the nursery framework. Utilizing a few electronic segments, for example, NodeMCU, Arduino, and various sensors associated over the web, the nursery parameters will be observed and controlled with the assistance of various electronic gadgets like advanced mobile phone, PC and so forth.

This proposed model of a Smart Greenhouse, encourages the nursery laborers to do the work in a nursery consequently without utilizing a lot of labor. Nursery, a shut situation shields the estates from cruel climate conditions and creature, bother assaults. The water system of savvy nursery is finished by the utilization of motorized dribble water system, which works on the parameter of soil dampness determined as ideal to the plants. In view of parameters from soil wellbeing card and information nitrate sensor legitimate measure of minerals can be applied by utilizing dribble water system procedures. Appropriate water the executives tanks are developed and they are loaded up with water subsequent to estimating the ebb and flow water level utilizing a ultrasonic sensor. Plants are given the fitting measure of light when it gets darker utilizing movable drove lights. The temperature and stickiness sensor are utilized to screen them.

We propose observing and controlling the nursery parts with the assistance of IOT. The ebb and flow model water tank for water system, water tank is definitely not an attainable arrangement our model uses dribble water system. The greater part of the current nurseries have inappropriate ventilation with absence of moistness control. We have proposed portable folds which work for controlling the moistness and temperature of the nursery.

I. OVERVIEW INTRODUCTION

Despite the fact that India gets bounty measure of precipitation and has various enormous waterway frameworks, yet 33% of the all out land for horticulture is joined by means of trench water system framework. Rest larger part of the segment depends on rainstorm or cylinder wells for the development of estate. There are places having overabundance water yet face issues of land saltiness, as a result of over water system and water-logging. The Greenhouse, being a shut casing structure defends the plants from overwhelming climate conditions, for example, sea tempests, Ultraviolet radiations, hailstorms, and bug and irritation assaults. Here, the water system of farming field is done with the assistance of a mechanized trickle water system process, which works as per the edge soil dampness that is being set so that an ideal measure of water is given to the plants all the time. In light of the information from soil wellbeing card, fitting measure of supplements, water and different imperatives might be provided by utilizing dribble water system methods. The plants are given the essential light wavelength through shining LED during the evening time. Inside the nursery, air stickiness and temperature are observed by the mugginess and temperature sensors separately, and can be controlled with the assistance of versatile rooftop folds which work consequently as it arrives at the limit temperature and moistness.

II. BACKGROUND

In India, farming is as yet completed in the customary manner and falls behind in using the cutting edge advances accessible to humankind. This model means to improve the previous agrarian practices by consolidating brilliant present day advancements for proficient and better creation of nursery. The undertaking gives a model of a shrewd nursery, which will help the ranchers in completing the homestead work consequently, in an unmanned way.

At present, there are various nursery models which work in a computerized way, yet a large portion of them need minor parameters, which affect the generation of nursery. With the assistance of the proposed model, these weaknesses can be proficiently survived, in this way improving the generation of the nursery. Our venture means to connect every one of these holes which will help in effective and less time of client being utilized in performing excess errands. At present, with the headway of innovation there have been a ton of enhancements in the nursery, yet at the same time there's plentiful space for further improvement to expand the productivity and generation pace of the nursery. There are nurseries which have dribble water system framework introduced in them, sensors that solitary gather information yet are not customized to play out any trigger-component based activities, so we expect to beat these deficiencies with the assistance of our proposed model. Generation rate is reasonable in the current nursery frameworks and can be helped with a lot of proficiency. Additionally, a large number of the nurseries have been subject to labor or human help with a few or the other way,

which can be changed into a totally robotized process by every minute of every day Real-time checking of the green house.

III. IMPORTANCE OF THE PROJECT

India being an agro-based nation, horticulture is as yet being completed in a conventional way, and along these lines falls behind in the joining of present day strategies and innovations. About 57% populace in India has been effectively occupied with farming and related exercises, which make for about 15% of the nation's GDP. Subsequently, it is significant in the present time for the included partners to abandon the old ordinary acts of farming and excel by modernizing the rural procedure with the assistance of better and proficient more up to date advances. The monetary commitment of rural exercises to the GDP of India is immovably diminishing with the nation's wide-extending development in economy, albeit an enormous number of individuals are still tirelessly occupied with the horticultural part.

IV. OBJECTIVES AND SCOPE OF THE PROJECT

The nursery ought to have the option to make various smaller scale atmospheres with exactness that favors custom condition for an assortment of plants. The nursery ought to have the option to monitor assets there by reducing expenses of the estate procedure and could prompt various development cycles in a limited capacity to focus time, expanding the efficiency and income for the customer.

- The robotized nursery in this venture is intended for residential utilize yet can likewise be scaled to be actualized for greater ranch. This venture centers around one plant as it were. Nonetheless, this reference can be changed into some other plant.
- The nursery can make a microclimate appropriate for the development of the plant, which was accomplished by controlling the variables, for example, mugginess, temperature, dampness, light source.

Other scopes of this project are:

- The robotized nursery in this venture is intended for residential utilize yet can likewise be scaled to be actualized for greater ranch. This venture centers around one plant as it were. Nonetheless, this reference can be changed into some other plant.
- The nursery can make a microclimate appropriate for the development of the plant, which was accomplished by controlling the variables, for example, mugginess, temperature, dampness, light source.

Summary

All the previously mentioned parts of present agrarian practices ought to be improved to get better return. Henceforth, we move towards a savvy nursery model where the plant is given a domain for its ideal development. Because of the shut structure of nursery, creepy crawlies and vermin can't enter inside, along these lines disposing of the necessity of bug sprays and pesticides. Sparkling LED lights are turned on at whatever point light force is low for photosynthesis, this guarantees quicker pace of development.

Our proposed model altogether improves Greenhouse efficiency which is fundamental for expanding gainfulness. Nursery profitability has been expanded by comprehension and anticipating crop/plant execution in an assortment of natural conditions which can be created by our proposed model. Existing nursery framework use a great deal of labor in controlling and checking the green house our proposed model has IoT computerization and gives constant information of the nursery condition for observing and control.

Flowchart

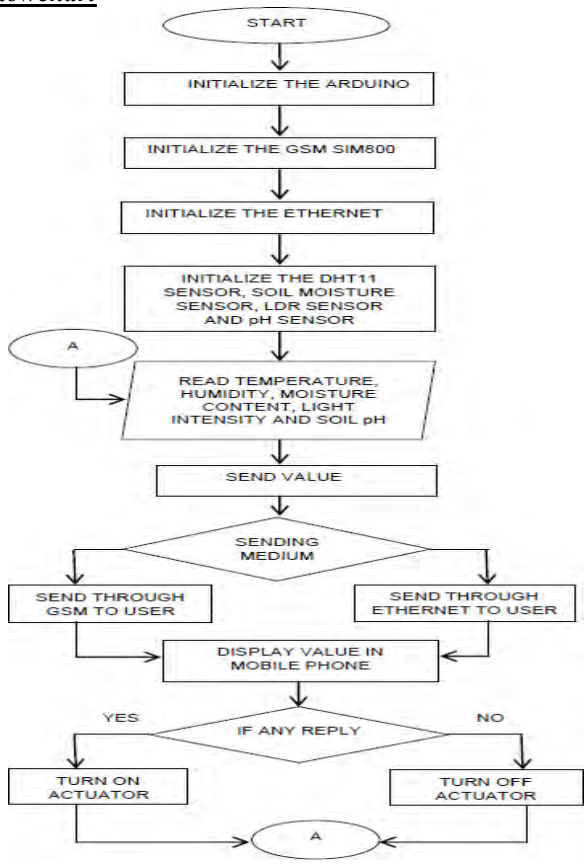


Figure No 3: Flowchart

V. RESULTS & DISCUSSION

IOT based systems are now a part of a variety of specializations from Agriculture monitoring to the Home Automation smart systems, which can be accessed from any part of the world, making it easy even for Farmers, doctors, working professionals to access the resources whenever

they require.
This is merely the beginning. Not long from today, IOT will be used to advance our understanding of wireless network through analysis of highly dense networks and huge datasets, help us invent new solutions and make smart systems more thorough and personal. IOT based Greenhouse systems can revolutionize agriculture sector and allow new paradigms in planning of architecture. Machines will run our homes more efficiently, make businesses more productive and help predict risks to society.

Actual Results

We were able to build a system for monitoring and controlling of the greenhouse system .Besides this, we were also able to add some basic features such as transferring that data to the device and showing the plants humidity ,temperature ,moisture on the device. The device which is required can be a monitor or mobile device which can be connected to Node MCU and can easily get the output.

Outputs: We were able to build a system for monitoring and controlling of the greenhouse system .Besides this, we were also able to add some basic features such as transferring that data to the device and showing the plants humidity ,temperature ,moisture on the device. The device which is required can be a monitor or mobile device which can be connected to Node MCU and can easily get the output.

Outcomes: The nursery ought to have the option to make numerous smaller scale atmospheres with exactness that favors custom condition for an assortment of plants. The nursery will be come up short evidence since every one of the parts will have the option to work autonomously so there is no single purpose of disappointment. The nursery ought to have the option to save assets in this way reducing expenses of the manor procedure and could prompt numerous development cycles in a limited capacity to focus time, expanding the efficiency and income for the customer.

Discussion:

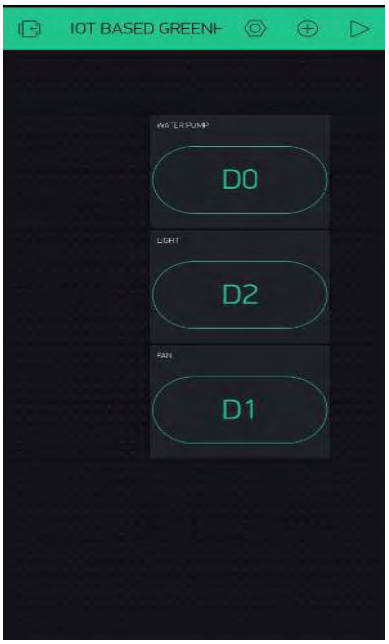
IOT based systems are now a part of a variety of specializations from Agriculture monitoring to the Home Automation smart systems, which can be accessed from any part of the world, making it easy even for Farmers, doctors, working professionals to access the resources whenever they require.
This is merely the beginning. Not long from today, IOT will be used to advance our understanding of wireless network through analysis of highly dense networks and huge datasets, help us invent new solutions and make smart systems more thorough and personal. IOT based Greenhouse systems can revolutionize agriculture sector and allow new paradigms in planning of architecture. Machines will run our homes more efficiently, make businesses more productive and help predict risks to society.

Future Scope

Intelligent agriculture is developing rapidly in recent years and it raises more attentions in both industrial and academic societies. This thesis take a full consideration of cost, practicability and other factors, combining the IOT with fuzzy control method ,using GPRS to remote control, designing an smart greenhouse monitoring system with better performance, simple structure and easy extensibility. There is also controlling action taken automatically that is greenhouse windows/ doors roll on/off based on the soil moisture levels. Thus, the system will help the farmers to avoid physical visit to the field, and increase the yield with the maintenance of précised parameters such as CO2, soil moisture, temperature, and light in the greenhouse with the help of IOT. The project is carried out with the help of IOT kit and internet connection.

Deployment

Deployment can be a major challenge in software which involves machine learning. Huge amounts of processing is required just to get it up and running. Data generated in the process is humongous as well. The ML model will be deployed using Google firebase or spring. The cloud platform will provide a base for future developments where flexibility, portability and reliability is required. The GUI by itself is not very difficult to deploy. We have multiple ways of offering interfaces, such as web interfaces and command line interfaces. It will expand to a desktop GUI in the future as well. The web interface will be deployed using Heroku or AWS, depending on the relative cost of hosting. The command line can be used as a windows executable file. Users can simply download it and run it natively on their windows PC



VI. CONCLUSION

We utilize the actuators (engines/moving parts) too, in this manner giving observing and information investigation as well as exact control for the brilliant nursery.

The results are analyzed for the greenhouse parameters such as CO2, soil moisture, temperature, and light for bell pepper plant with the help of graphical representation based on the practical values taken by the IoT kit. The comparative result shows the effectiveness of the proposed work.

References:

[1]. ShreyasBhujbal, Yash Deshpande,Arpit Gupta, OjasBhelsekar "IOT Based Smart Greenhouse" IJIRSET, Vol. 7, Issue 1, Jan 2018

[2]. Prof. C.R. Dongarsane, Patil Pranav Balasaheb, PatilNileshRangrao, PatilPranit Ramesh "Green House Automation Using IoT" IRJET, Vol. 4, Issue 1, Jan 2017

[3]. VeenaPatil, Sonali B Patil, Swetha S Gogi, Satappa M Paramagond, Sidharta A Vardhamana "WSN BASED GREEN HOUSE OR IRRIGATION MONITORING AND CONTROL SYSTEM USING SMART PHONE" IRJET, Vol. 4, Issue 5, May 2017.

[4]. Ravi Kishore Kodali, Vishal Jain, SumitKaragwal"IoT based Smart Greenhouse" IEEE, 2016

[5]. Imran Bin Jafar, Kanij Raihana, SujanBhowmik, Shifur Rahman Shakil "Wireless monitoring system and controlling Software for Smart Greenhouse Management" IEEE, 2014.

IOT Based Grain Moisture Monitoring

Anshu Singh, Hetal Amrutia, Manish Singh, Yash Shukla

Information Technology Thakur College of Engineering and Technology Mumbai, India

s.anshu4855@gmail.com, hetal.amrutia@thakureducation.org, ms292735@gmail.com, vashshukla048@gmail.com

Abstract— **Moisture is one of the important parameter which is always considered when deciding the quality and price of grain, at the stage of harvesting, storage, processing and marketing. Grain having excess moisture content, if stored for long duration, will spoil due to insect/fungus infestation. The types of moisture meters available in the market are with the lookup tables which cause inconvenience when carrying out measurements. These previously existing moisture meters are inaccurate sometimes due to their way of measuring the moisture of the grains with respect to a constant room temperature and the user also has to check the lookup table for checking the moisture index which is way too hectic for the user. So to overcome this problem we have developed IOT based Grain Moisture Meter which due to its novel design gives moisture (%) on an android application and also on a web application. The moisture meter will provide an analysis report of the grain storage.**

I. INTRODUCTION

Accurate measurement of grain moisture content is important during harvesting, drying, storage and processing operations. The accuracy of the moisture meters in measuring the moisture content of the grains is very important to ensure fair trade between farmers and buyers. Electrical moisture meters cannot measure the moisture content of a commodity accurately due to a variety of reasons. The parameters of measurements of these moisture meters are affected not only by the moisture content but also by the sample size, shape, temperature and physical and chemical constitution of the commodity.

It is therefore essential to test and calibrate the relationship between electrical reading and the actual moisture content, determined by standard oven method. An air oven method was adopted as the standard method in moisture determination of grains and seeds (ASAE, 2003) and method outlined by the International Organization for Standard (ISO) (Anon 2001) have been used to standardize simplified oven methods that are regarded as practical working methods. During procurement there is now available an inexpensive method of sufficient accuracy used generally by farmers for determining grain or/and forage moisture content. In most cases, the farmer relies upon his experience to estimate the moisture content. For grain, this might entail biting, rattling, feeling, and observation but officially there is always need for rapid moisture measurement and so the electrical instruments have been developed. As long as the grain moisture-sample level is not too high, such instruments have been relatively satisfactory. At moisture content above 20-25%, however the consistency and reliability of measurement with existing moisture meter begins to deteriorate. Over the past 50 years, numerous instruments have been developed for measuring grain moisture content by electrical means. Instruments that sensed the dielectric properties then became more popular because dry outer layer of the kernels did not degrade their reliability as much. The dielectric properties of grains were also studied in connection with radiofrequency dielectric heating for grain drying. These electric properties of grain have long been utilized by electric and electronic grain moisture meters, which sense the electric conductivity or dielectric properties of grain and are calibrated to read moisture content.

The knowledge about electric properties of grain is also useful in designing of equipment for electro treatment of grain, i.e., dielectric drying, electrostatic cleaning and sorting, desensitization, and also electrical stimulation (Datta et al, 1995). Consequently, there is a need for development of instruments with better accuracy, consistency and reliability for rapidly measuring the moisture content of high moisture grain.

II. LITERATURE SURVEY

A. Research Gap

Focuses more on micro organisms, Insects and Rodents rather than humidity and temperature control. Uses canola/rapeseed storage time chart that predicts grain quality only for 5 months. No focus on air conditioners/room chillers in room to maintain the cooling temperature for grains stored in warehouse. Uses capacitance principal which is the one of the oldest methods. Every time the user will need to check the display to get the moisture level. Not giving 100% accurate output and the developed system is working satisfactorily. This paper only tell about wheat grain and no other grains in the market. Uses only direct method to prevent moisture content of wheat and having approx. estimation of moisture. It does not show how the grains can be protected from high temperature and humidity in the warehouse. It only tells that how wheat moisture can be maintained after harvesting. This paper tells more about measurement of water in the grains, sand rather than showing humidity affecting it. Samples with very high moisture content (say 60% and above) the insertion loss is too great to be measured correctly. This method has the drawback, though, that a single measurement can take up considerable time. Suggested to use hermetic bags instead of non-hermetic bags which is expensive and challenging. The PHL meter, John Deere meter and GAC 2100 meter was suggested to use instead of oven dry method which is mostly used nowadays. Conducted measurement on maize grain and had to check the moisture every month manually. For

measuring the grains different meters are used from which the work is done manually which is time consuming and hectic for humans to do the work. Price of moisture meters is expensive and for unique grains different meters are required such as:

- Resistance type Indosaw (2000)
- Resistance type osaw (40471)
- Capacitance type Indosaw (DM-06)

III. OBJECTIVE AND SCOPE OF THE PROJECT

1. When Quality grain production:

Crop quality is determined by soil nutrition, crop management, crop handling, and weather. The two main obstacles to producing high quality organic grain are insufficient soil nitrogen and the farmer's lack of knowledge and experience. Many new organic farmers have just moved to the country, and bought depleted or marginal land. Their poor soil and lack of experience lead to a poor quality crop. Sometimes farmers are 'organic by neglect' in that they don't use any inputs -- neither the prohibited pesticides and synthetic fertilizers, nor useful organic amendments. In contrast, successful organic farmers are 'organic by design'.

To produce high quality food and be 'organic by design', walk the fields and pay attention to details. Learn to recognize weeds, crop quality and changes in the soil. Take the effort to learn about the farm, and to develop a good understanding of soil, plants and animals. Learn from experience, from talking to other farmers, reading books and going to workshops. A commitment to quality improvement and an investment of time and dedication will lead to better crops.

Being 'organic by design' requires good planning and problem prevention. Plan your crop rotation three years ahead, but check the growing conditions and adapt the rotation as needed. Off-farm inputs are used if needed, including soil amendments. A wise selection of high-quality tools is purchased for the farm.

2. Grain storage management:

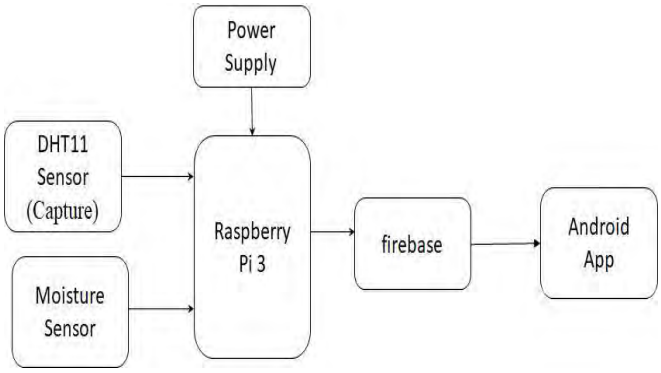
As our project is going to give the moisture, room temperature and humidity in the warehouse due to which the warehouse manager will get to know whether if the room temperature or moisture increases or decreases then what measure is to be taken. Accordingly the warehouse manager will shift the crops or will keep the doors open to maintain the room temperature or increase or decrease the moisture of the grains so that the quality of the grain remains the same it does not gets degraded. It would help the warehouse manager to manage and store the grains accordingly in the suitable condition.

3. Analytics of grain moisture:

The moisture, room temperature and humidity of the warehouse or the place where grains are stored this data can be stored to the database which would help the farmers to get a high-quality grain production.

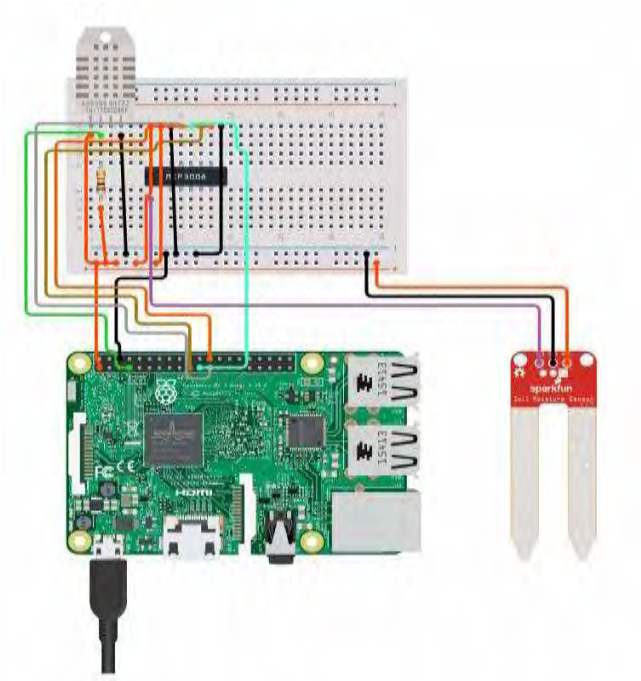
IV. PROPOSED METHODOLOGY

A. Block Diagram:



The dht 11 and moisture sensors are connected to the raspberry pi 3 B+. DHT 11 sensor gives the readings of room temperature and Humidity of the Warehouse through GPIO pins. Moisture Sensor gives the moisture readings of the grains in a sack. The code which we will implement will capture the data in json format and push online to the firebase database. There would be a android application where we will retrieve this Json data and process it to show the quality of the grain and the condition of the grain. The data in the firebase will help us to analyze whether the crop is in good condition or not and we will get to know in which period that is in which season the farmer has got the quality production of food grains. The android app will give the notification to the farmer to the farmers that if the quality is degraded then what is the solution. The app will also help in inventory management that is how many sacks of grains were sold and how many sacks were of which quality.

B. Circuit Diagram

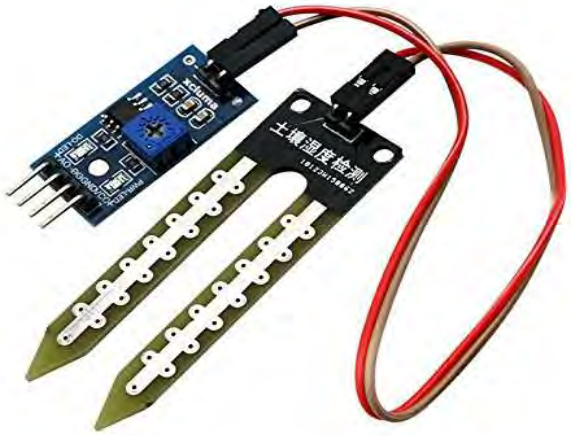


Here we have connected the soil moisture sensor to the raspberry pi using MCP3008 which is an analog to digital converter as the raspberry pi does not consist of a analog pins. Dht 11 is directly connected to raspberry pi at GPIO pin 27. MCP3008 gets the analog signals from the Moisture sensor and converts them into digital signals and passes it to raspberry pi.

All sensors used in our system are RoHS certified, according to which the sensors are made up of environment friendly materials. The sensors used are:

a. DHT11:
In our proposed system this sensor is playing an important role to get the current room temperature, which will help to know whether the grains are stored in the favorable weather conditions. The dht 11 provides the data in digital form so it is easy to connect to our system on chip (Raspberry Pi 3B+). The information is retrieved from sensor via using Adafruit_GPIO.SPI in Python, this library helps to retrieve and process the data. This sensor also gives the humidity of the warehouse. After getting data from the sensor it is stored in the json file, easy to access and light weight database.

b. Moisture Sensor:
This sensor is the backbone of the system it helps to retrieve the actual moisture of the grains which are in the grains. The sensor provides data in the analog form which is not compatible to our system on chip as the raspberry pi consists of all digital pin. To make our sensor compatible i.e. to get the data in digital form we are using Analog to Digital Convertor MCP3008. The MCP3008 helps to convert the analog data into digital data which helps our SOC to read the data easily. To get this data and process it accurately we use spidev library provided by the sensor manufacturer itself. After getting data from the sensor it is stored in the json file, easy to access and light weight database. The stored data can be accessed via a web portal. The web portal also provides the live graph of the moisture sensed via sensors.

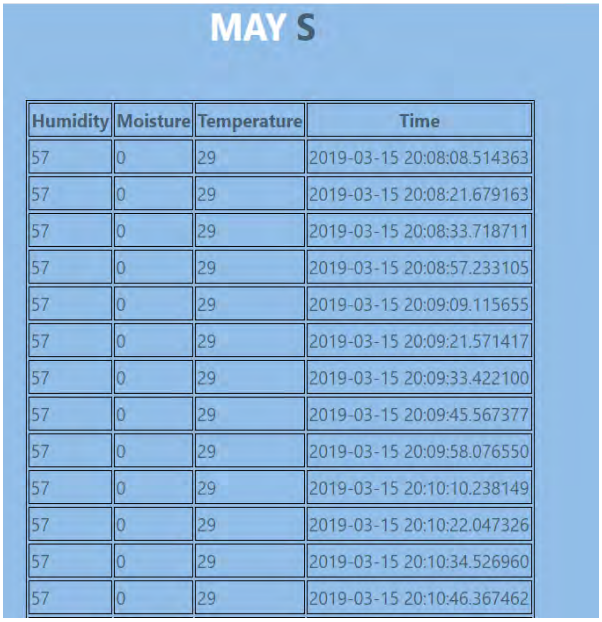


c. MCP3008:
It is a analog to digital convertor which helps to get the analog data of moisture sensor to the raspberry pi in digital form.

V. ACTUAL RESULTS

a. Outputs (Output of the project only)

Successfully implemented the Moisture meter for grains.



MAY S

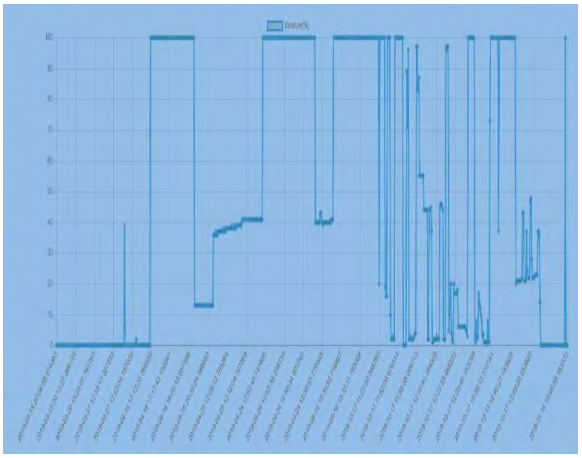
Humidity	Moisture	Temperature	Time
57	0	29	2019-03-15 20:08:08.514363
57	0	29	2019-03-15 20:08:21.679163
57	0	29	2019-03-15 20:08:33.718711
57	0	29	2019-03-15 20:08:57.233105
57	0	29	2019-03-15 20:09:09.115655
57	0	29	2019-03-15 20:09:21.571417
57	0	29	2019-03-15 20:09:33.422100
57	0	29	2019-03-15 20:09:45.567377
57	0	29	2019-03-15 20:09:58.076550
57	0	29	2019-03-15 20:10:10.238149
57	0	29	2019-03-15 20:10:22.047326
57	0	29	2019-03-15 20:10:34.526960
57	0	29	2019-03-15 20:10:46.367462

The DHT11 Sensor detects the level of the humidity and temperature of warehouse and moisture sensor measure the level of grains moisture and alerts the system, of the increase in the level. The result would be displayed on the webpage. This helps the concerned authority to carry on a require action.

b. Outcomes

The following are the outcomes were observed as a result of implementing the product:

- Temperature and Humidity level detection of warehouse.
- Moisture level of grain dictated in real time wirelessly.
- Transmit the information wirelessly to

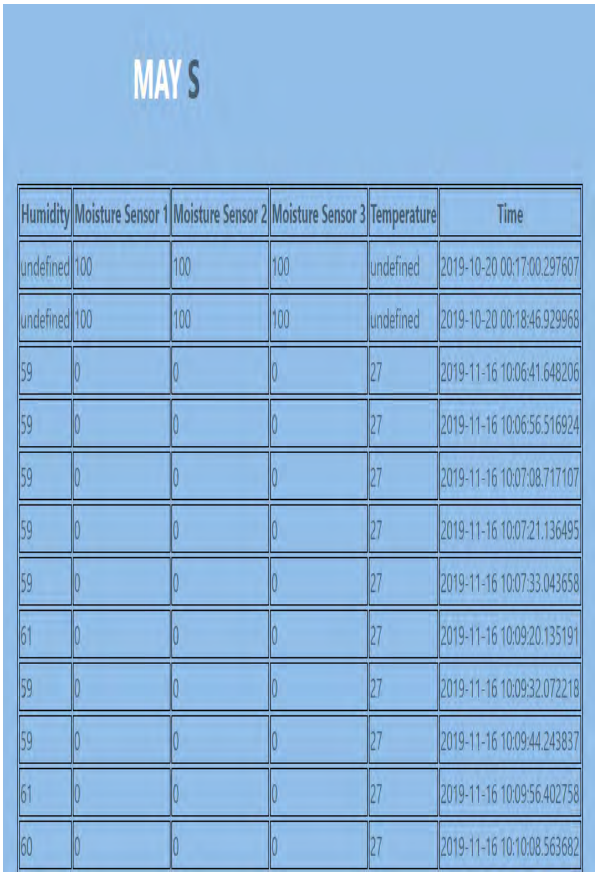


Webpage.

- The data can be accessed anytime and from anywhere.
- The real-time data transmission and access.
- Avoids the Grain Wastage.

c. Discussion of the results

This IoT based wireless grain moisture and warehouse temperature detection system, which is very useful for Warehouse and even the rural areas as equally. In the Warehouse lack of the grain moisture level management has been leading to wastage of grains. In order to eradicate this kind of wastage of grain, this system is developed. Hence the monitoring will widely help them take the necessary precautions and prevent the grain from degrading in warehouse. IT has been observed that tons of grain get degraded/destroyed because of increase/decrease of moisture level in warehouse. The system is designed to solve this issue and will provide complete details of the grain moisture in the warehouse. The concerned authority can access the information from anywhere and anytime to get the details. Accordingly, they can take the decision on this immediately. In addition to the added advantages of being able to monitor the grain moisture and warehouse temperature wirelessly, a feature has been incorporated where the all detail about grain can be access through Android Application. This particular feature further enhances the capability that this particular product or project brings to the table. This adds to the existing feature of being able to monitor, but when the



MAY S

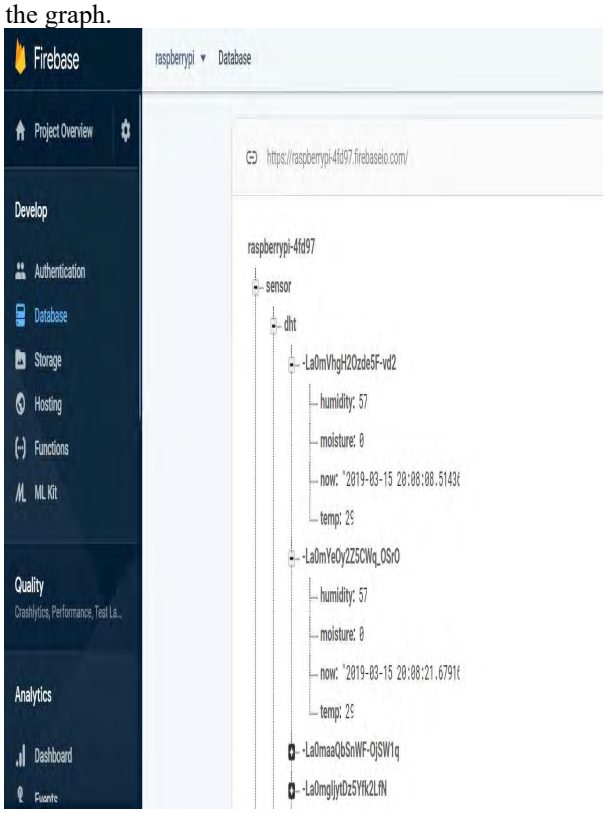
Humidity	Moisture Sensor 1	Moisture Sensor 2	Moisture Sensor 3	Temperature	Time
undefined	100	100	100	undefined	2019-10-20 00:17:00.297607
undefined	100	100	100	undefined	2019-10-20 00:18:46.929968
59	0	0	0	27	2019-11-16 10:06:41.648206
59	0	0	0	27	2019-11-16 10:06:56.516924
59	0	0	0	27	2019-11-16 10:07:08.717107
59	0	0	0	27	2019-11-16 10:07:21.136495
59	0	0	0	27	2019-11-16 10:07:33.043658
61	0	0	0	27	2019-11-16 10:09:20.135191
59	0	0	0	27	2019-11-16 10:09:32.072218
59	0	0	0	27	2019-11-16 10:09:44.243837
61	0	0	0	27	2019-11-16 10:09:56.402758
60	0	0	0	27	2019-11-16 10:10:08.563682

authorities are away, it gives them an added option of continuous monitoring.

d. Database Outcome:

Here in our system we have used the google firebase database. Google Realtime Firebase database stores the live data. It stores the data in JSON format; which is a light weight file and can be accessed easily. JSON data can be converted to multiple file formats according to the usage and convert that JSON data to csv file format

which is used by python to provide analytics report or the graph.



VI. FUTURE SCOPE

In future we will be adding inventory management module which will be helping them to know how many sacks of grains were sold to whom and at what time. The warehouse manager could also see the monthly or yearly billing of the grains which were sold, which will be helpful to calculate and order grains from the farmers at value price.

VII. CONCLUSION

Apparently, a systematic study has never been conducted over a wide range of moisture in grains by conditioning or drying to the range of different grain moisture meters and various characteristics of grain that contribute to variability in dielectric properties. In addition, there should be more study conducted on the how the physical properties of grains moisture content during procurement, storage and processing. The best dielectric properties and best frequency ranges should be selected while designing the instrument for quick moisture determination. The proper selection and use of circuits are as important as to design the sample holder capacity for different commodities. The other important requirements for grain moisture measuring instrument include the ease of use, accuracy, consistency, precision, reliability and light weight so that one can easily carry the instrument from heap to heap during procurement season. Developed a moisture meter for grains which is capable of sensing room temperature and humidity of warehouse as well as moisture of the food grains. Collecting data and research can be done from this proposed system and can protect the grains from getting rancid or spoiled.

References

[1] Author Name Kavya P, Pallavi K N, Shwetha M N, Swetha K, Mrs. Jayasri B S, “Use of Smart Sensor &IoT to Monitor the Preservation of Food Grains at Warehouse”,IJRTI,2017.

[2]Author Name Miss. Rachana Gumma, Miss. Rutuja Magar, Miss. Anuradha Dhavan, ProfJagtap S. C, “Development of Grain Moisture meter with Moisture and Price Display”,IJTSRD, Mar-Apr 2018,volume2.

[3]Author Name Babankumar S. Bansod, Sucheta Gupta, “Calibration of moisture meter for cereal (wheat) grain, its evaluation & error analysis at 230KHz based on microwave oven techniques”,IEEE,2009.

[4] Author Name Hristomir Yordanov, Vasil Kumanov, Nickola Kumanov, Lyubomir Urshev, Boris Vichev, “Calibration Techniques for Microwave Moisture Meters”,IEEE,2016.

[5]Author Name Shekinat K. Ajao,Adeola A. Ala , Kehinde O. Popoola , Mobolaji O. Omobowale,George P. Opit, “Comparative evaluation of novel low-cost moisture meters suitable for grain moisture measurement”,IEEE,October, 2018.

[6] Author Name Usman Ahmed, Mahesh Kumar, Salam din, BVC Mahajan ,M S Alam,“Grain Moisture Measuring Techniques - A Review”,CIGR ,PAGENO:13-18,March 2015.

[7] Nikkila, R., Seilonen, I., Koskinen, K. 2010. „„Software Architecture for Farm Management Information Systems in Precision Agriculture.“ Comput. Electron. Agric. 70 (2), 328-336.

[8] Alexandros Kaloxylos, J Wolfert, Tim Verwaart, Carlos MaestreTerol, Christopher Brewster, RobbertRobbemond and Harald Sundmaker. „„The Use of Future Internet Technologies in the Agriculture and Food Sectors: Integrating the Supply Chain“ in 6th International Conference on Information and Communication Technologies in Agriculture, Food and Environment. pp. 51-60

[9]Kevin Ashton, „„That Internet of Things thing“ RFID Journal, It can be accessed at :http://www.rfidjournal.com/articles/view?4986.

[10]G. Tripathi, A.J. Jara, „„A survey of Internet-of Things: Future Vision, Architecture, Challenges and Services in Internet of Things (WFIoT), 2014

[11]Gartner, Inc. It can be accessed at: http://www.gartner.com/newsroom/id/2905717.

[12] Malik Tubaishat, Sanjay Kumar Madria „„Sensor networks: An Overview“, IEEE Potentials 05/2003.

[13] Juan Felipe Corso Arias., Yeison Julian Camargo Barajas., Juan Leonardo Ramirez Lopez., „„Wireless Sensor System According to the Concept of Internet of Things International Journal of Advanced Computer Science and Information Technology Volume 3, Issue 3, 2014, ISSN: 2296-1739

[14]Tadele Tefera, Fred Kanampiu, Hugo De Groote, Jon Hellin, Stephen Mugo, Simon Kimenju, YosephBeyene, Prasanna M. Boddupalli, Bekele Shiferaw, Marianne Banziger. „„The Metal Silo: An effective grain storage technology for reducing postharvest insect and pathogen losses in maize while improving smallholder farmers’ food security in developing countries“, The International Maize and Wheat Improvement Center (CIMMYT), Volume 30, Issue 3, March 2011.

[15] Grant R. Singleton. „„ Impacts of rodents on rice production in Asia.“ IRRI Discussion Paper Series No. 45, 30 pp. (International Rice Research Institute: Los Banos, Philippines.)

[16] Zeleny, L. (1960). ‘Moisture measurement in the grain industry,’ Cereal Science Today, Vol 5, pp 130-13

[17] Hunt and Haward, W. (1965). Humidity and Temperature Vol. II, Reinhold Publishing Corp., New York, p 123.

[18] Gough, M.C. (1983). “Moisture Meter Calibration: A Practical Guide”, Tropical Stored Product Information, Vol. 46, pp 17-24.

[19] Masson, I. (1911). ‘The use of calcium carbide for determining moisture’, Chem. News, Vol. 103, pp 37-38.

[20] Fisher, K. (1935). ‘A new method for the analytical determination of the water content of liquids and solids’, Angew. Chem. Vol. 48, pp 394-396.

Encryption Environment for IoT Data Security

Mr. Ankit Sharma, Mr. Kishan Sharma, Mr. Yash Tandel, Mrs. Vandana Munde
Information Technology Thakur College of Engineering and Technology
Mumbai, India
aravsharma3456@gmail.com, kishansharma467@gmai, tandelyash874@gmail.com, vrsakharkar@gmail.com

Abstract- — **In Internet of Things (IoT),lot of information is being shared between the devices like smart houses,self driving cars etc. But the information is not protected as the Hackers can hack our system and also steal or change our information. In such circumstance there is a need for protecting the data using cryptographic algorithms. With the boom of Internet of Things, the embedded devices are being able to connect to the internet and intercommunicate with each other. Internets of Things find many applications in various fields but the emergence of IoT have put the embedded device communication and interaction at risk. There are various security conditions that need to be satisfied to be able to securely connect the embedded device to the internet. The information is not protected as the Hackers can hack our system and also steal or change our information. In such circumstance there is a need for protecting the data using cryptographic algorithms. In this Paper a strong Encryption Cipher Called PRESENT which is an Ultra-Lightweight Symmetric Block Cipher is discussed as an application of PRESENT cipher,Text message encryption is considered. Elliptic Curve Diffie Hellman (EC-DH) Algorithm has emerged as an attractive and effective public key cryptosystem. Elliptic curves are widely used in various key exchange techniques that include the Diffie-Hellman Key agreement scheme. The proposed work also involves analysis of power, performance, area, and their comparisons thereof. The comparison is based on metrics obtained, after implementing the algorithms in synopsys using 90 nm UMC Faraday library. The ECDH algorithm is found to be better than others as far as power and area are concerned. The results of IOT failures can be severe, therefore the study and research in security issues in the IOT is of extreme significance. The main objective of IOT security is to preserve privacy, confidentiality, ensure the security of the users, data and devices of IOT .Thus research in IOT ecosystem has recently been gaining much momentum .Traditionally the data was compromised during the time of transmission, facing problems like confidentiality, Integrity etc. As a traditional approach, only present algorithm was used which is a light weight algorithm**

I. OVERVIEW INTRODUCTION

The Internet of Things (IoT) brings a new set of security concerns. Unlike VPN encryption, which safeguards networks through an encrypted and anonymous tunnel, IoT devices must be inbuilt with their own strong security and encryption standards. VPN allocates an isolates space on the network. However inside VPN space, all of the network nodes can be accessed by any participant. When IoT devices are out there, a VPN has many open doors Securing the Internet of Things is critical not only for the integrity of data streams and software within each of the IoT application, but it is also for the integrity of the enterprise resources tied into those applications. IoT security is a complex problem, which requires a systematic approach for understanding possible threats and corresponding mitigation methods. In Internet of Things (IoT),lot of information is being shared between the devices like smart houses,self driving cars etc. But the information is not protected as the Hackers can hack our system and also steal or change our information.In such circumstance there is a need for protecting the data using cryptographic algorithms. In this Paper a strong Encryption Cipher Called PRESENT which is an Ultra-Lightweight Symmetric Block Cipher is discussed as an application of PRESENT cipher,Text message encryption is considered. In today's technology, there is an increasing number of electronics applications require secure communication for example the IoT devices. Elliptic Curve Diffie Hellman(EC-DH) Algorithm has emerged as an attractive and effective public-key cryptosystem. Elliptic curves algorithm are widely used in various key exchange techniques that include the Diffie-Hellman Key agreement scheme. When contrasted with conventional cryptosystems like RSA,ECC offers equivalent security with smaller key sizes, which results in lower power consumption, speedier calculations, and also lower memory and transmission capacity (bandwidth) reserve.What we are doing is that we are implementing present lightweight algorithm along with Deffie Hellman elliptical curve algorithm, which will provide more security to transfer the data.

Phase 1: deals with the deep study of the existing systems,by analysing the previously available research papers and technology.
Phase 2:The result of the literature survey and analysis based on various algorithms using previously available data on IoT security.
Phase 3:The parameters that are major reasons for the data security and attacks are detected and analysed to produce the design of the system.
Phase 4:implementation of the design using technologies planned in the above phases.
Phase 5: Testing and implementation of the software using the tools and the technology.

II. BACKGROUND

The main objective of this project is to develop a secure path of transmission of data between two devices that will secure the connection and the data which is been sent to the other iot devices, so that no other device can interfere in the transmission. This will be done by using the encryption and the decryption methods to ensure

confidentiality, integrity and availability of data. The algorithm used for encryption is Present algorithm and for decryption is Diffie Hellman elliptical curve..

Importance of the project

To Secure Device Hardware Accessibility :

Software security has proven inadequate to thwart the known threats. A networked device which lacks adequate hardware security can be easily hacked by an end-user. So, can you overcome this security flaw and secure your IoT devices against end-users? By encrypting and protecting configuration bit streams.

To Secure Port Communication :

Communication to and from the device needs to be secured using encrypted communication. One way to ensure this is avoiding the use of insecure encryption algorithm. Using weak encryption algorithms creates a false sense of security. We always think that our encrypted data will never be decrypted and stay hidden as long as we want. However, that will only be if you use solid encryption algorithms or follow secure encryption processes.

To Use Encryptions:

Encryption can help to prevent unauthorized access to data and devices. Also part of a complete security management process, encryption uses standard cryptographic algorithms to secure data at rest and in transit between IoT edge devices and back-end systems. This helps maintain data integrity and prevent data sniffing by hackers.

User Authentication:

All communication with your IoT devices should be authenticated using strong passwords, authentication protocols or time-based authentication tokens.

To Ensure Protection Against Phishing and Malwares :

Antivirus software which are useful can provide a critical layer of protection against attacks. They limit communication to only to the known, trusted hosts, blocking hackers before they can launch an attack.

Employ Middleware Hardware Physical Security:

A category of software, Middleware Hardware Physical Security provides a variety of platform and applications which are designed to integrate multiple unconnected security applications and devices and control them through one comprehensive user interface. By controlling access to hardware and software, physical security eliminates the possibility of any other device or user to damage or steal devices or data. Moreover, it limits physical interaction with equipment and implements systems to ensure that equipment is safe from environmental threats and to promote physical security.

Objectives and Scope of the project

III. OBJECTIVE

- Ensuring security of personal data on public databases: Interviews with our respondents suggested that there was a lot of concern with respect to two important issues. The first issue is unnecessary storage of private data on public forums. Many of the IoT devices and applications constantly stream data to other devices and servers. And this data never gets deleted. Rather it is an aggregated and used for trend analysis or evaluation of browsing histories. Second issue is indefinite availability of search histories.

- Maximize data integrity: Data integrity emerges as a fundamental issue in IoT. This is because there is a constant transmission of data from one device to the other. And there is an increased change of unauthorized changes and the accuracy of the data that is received. It is therefore paramount that technologies that enhance integrity be integrated into IoT. One of the respondents in our study noted:

- Maximize security awareness of IoT: While security awareness has always been a vital issue to consider, its importance is elevated in an IoT environment. This objective dealswith the increasing awareness of IoT and educating the consumers on IoT security. In a survey conducted by AT&T, it has been concluded that only 10 percent of respondents are fully confident with that their devices are fully secure. About 68 percent of respondents say that their companies are planning to invest IOT security in 2016.Many organizations deploy IOT devices, but in some of them the IOT devices are deployed without proper security measures.

- Enhanced protection of personal information on devices: This objective emphases the importance of placing reasonable limits on the collection and retention of data by IoT. When a large amount of data is stored, it is more likely at the risk of being hacked by the intruders, thus increasing the potential harm to the personal information on the devices. Scope: To be able to securely send the data hashing can also be used .

- MD5 is a hashing algorithm which outputs a message digest which is of fixed length derived from an input of arbitrary length.

- The designed algorithm will have various features that can be used to satisfy the security concern in IoT. The proposed research about present algorithm will allow the use of MD5 for securing the embedded device from attacks when connected to the internet to become a part of the Internet of Things.

Feasibility study

Economic

The project relies on open source software, such as Python, MATLAB. The functionality provided by these software's is enough to construct a product and capable of making project.

Technical

The capabilities required for the project are fairly in the feasible range. Most of the planned feature hinge around linking Databases. A GUI is required, which will be done with the pyGUI framework. The GUI will be kept simple and minimalist. We also have a feature for users to set personal data, preferences.

Operational

On an operational level, the software will need to run intensive processing tasks while it's in its learning phase. The software will be deployed to the IoT devices which eill help in encrypting and decrypting the data which is been send from the devices. It will need internet to operate.

Legal

Since the project deals with a sensitive topic with potential for major harm, it will carry a disclaimer for improper usage. The software is not meant to be use for any illegal activity. However, the potential for misuse, whether intentional or not, still exists.

IV. METHODOLOGY

- From Literature survey, we understood how the data is attacked in IOT device transfer and also how it can be prevented.
- From Gap analysis, we understood that the prevention techniques of data been accessed by other devices during transmission but all this techniques are not fully supported and they sometimes cannot be trusted.
- From problem definition, we understood that as of now there is no permanent solution for prevention of data been accessed by other user.
- From proposed solution, we decided to build an application that will stop the other device to gain access to the transmission device so that no data is leaked.
- From result, we understood how our system will analyze the data, encrypt the data, use public and private key for encryption and decryption and simultaneously get the data in the normal form at the other end.

Product Backlog or Sprint backlog

The backlog of the project is given as follows:

1. All the coding will be done in Python and C.

2. Simulation will be used to show the real time application of project.

3. There will a shared secret key exchange between sender and reciever using ellptic curve diffie Hellman key exchange algorithm

4. Then the input from the sender will be encrypted using the shared secret key and present algorithm and send to the receiver through the channel.

5.The encrypted data will be received by the receiver and decoded using the decoder and the shared secret key.

6. Design
   - Creating simulation environment
   - GUI building
   - Modeling
   - Feasibility analysis

Project Planning

IT projects require resources in terms of money, time, human resources, infrastructure and technology, both hardware and software. Resources are not just a mean, but also an approximation of constraints. Project planning is essential to managing the scope, schedule and budget of the project. For this, we used tools such as Matlab,Turbo C,Sublime text ,python ide, MS PowerPoint, online MS project as well as various modeling tools.

We divided our project into various phases and sub-phases, and allocated date ranges from a week to 3 weeks to every sub-phase. This was done using the timeline chart feature of online MS project. We then used a Gantt chart to model schedule dependencies and fine tune the scheduling. The Gantt chart was created using the Gantt chart feature available in MS Excel. Using these two, we further planned our sub-phases.
The modeling included charts and diagrams such as UML diagram, GUI modeling, dependencies and other diagrams. These were done using free tools like draw.io. Drawing these models helped to incisively assess our requirements and features.we analyzed various prevalent research papers pertaining to our project as a part of the literature survey. The project aims at researching about the encryption algorithm in IoT network. Tapplication will be able to help organizations secure their SQL database.

Our literature survey was focused on the capabilities of the choices we made for the core part of the project. We have divided the project into 2 phases.

- Phase I – Phase 1 will focus on forming a base for the project. With thorough planning, analysis and design, we will ensure that the actual implementation is smoother.
- Phase II – Phase II will focus on fine tuning the project and ensuring that the intended features will work as we wanted them to.

Scheduling

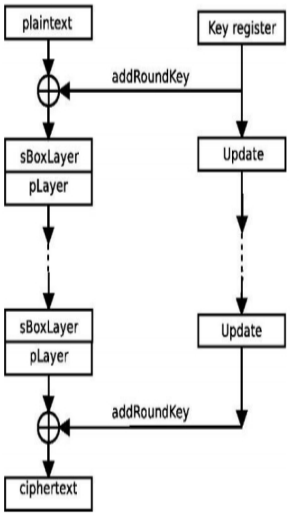


Figure 1: Gantt Chart

Design and Implementation

Flowchart

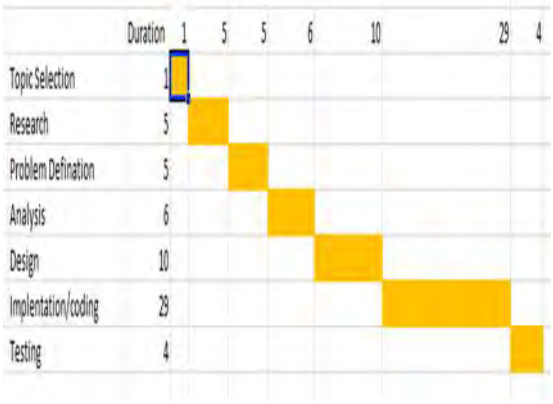


Figure 3: Flowchart

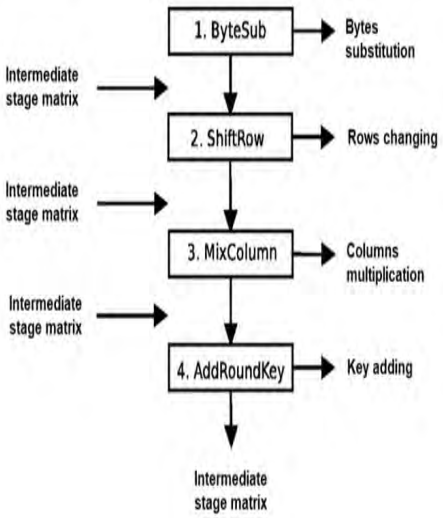


Figure 2: Block ciphers basic operation

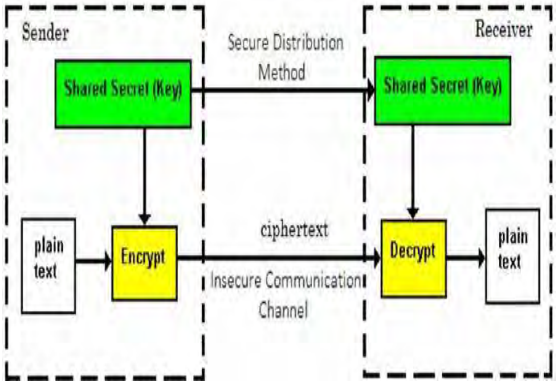


Figure 4: GUI

V. RESULTS & DISCUSSION

Actual Results

A system is proposed for secure communication betwseen two devices over the internet by using the encryption and decryption.
Using present algorithm we will protect the data of the user and confidential data in IoT devices the data will be encrypted using public key encryption.

a. Outputs:

Sprint 1: Literature survey and deep study of previously available system provided the outcome as the requirements of this project

Sprint 2: Documentation and formulation of the design.The ultimate design of the system is the overall output.

Sprint 3: The design is being implemented with GUI and the hardware

b. Outcomes:

The outcome of the above sprints are, sprint 1 provided the key findings and research gaps that can become requirements for the projects. Sprint 2 contributed by finalization of the system and services to be provided by the project. Sprint 2 also provided the design of the overall system. Sprint 3 provides output as 50% implementation of the system.

Outcomes:

This algorithm will provide security to IOT devices by using deffie hellman elliptical curve algorithm for key exchange and present algorithm for encryption and decryption.
Using present algorithm we will protect the data of the user and confidentiality of data in IOT devices will be encrypted using public key cryptosystem.

Future Scope

For future research, the implementation of the algorithm on hardware and software in various computation and network environment is under consideration. We will also use the advance hashing techniques like MD5 hashing algorithm which outputs a message digest which is of fixed length derived from an input of arbitrary length Moreover, the algorithm can be optimized in order to enhance the performance according to different hardware platforms. The scalability of algorithm can be exploited for better security and performance by changing the number of rounds or the architecture to support different key length.

Testing

Once the program code is designed and implemented, some testing technique will be used to ensure the program function correctly.

Unit Testing
Individual modules will be tested. The project is made up of smaller parts that can tested individually against the specification and design to confirm their correct operation.

Integration Testing
Several units will be tested together to see how they interact and to confirm whether their overall function is performed correctly. This testing will apply to each of the main section of code; the use interface, data processing, etc.

System Testing
The entire system is tested against the specification to check it meets the project's requirements.

.

VI. Deployment

Deployment can be a major challenge in software which involves networking.
The GUI by itself is not very difficult to deploy. We have multiple ways of offering interfaces, such as web interfaces and command line interfaces. It will expand to a desktop GUI in the future as well.
The web interface will be deployed using heroku or AWS, depending on the relative cost of hosting.
The command line can be used as a windows executable file. Users can simply download it and run it natively on their windows PCs.

VII. Conclusion

The internet of things is growing exponentially. It has spread from the power grid to smart refrigerators at home and at industrial levels by monitoring production line efficiency. The devices that connect to the internet must be encrypted because of the personal and business intelligence data they transmit. IoT users and security managers need to be attentive to security and to the ways encryption can both power and protect the next generation of secure networks and devices. The best protection available right now is encryption. The security requirements are similar to those of earlier generations of networked devices, embedded devices or security tokenization but need to be combined together strategically to match each specific use case. The security design of an IoT node can greatly benefit from the experiences gathered in the fields of platform, network security, and highly sensitive applications such as payment or financial transaction, government related application, content protection. The solution must be adapted to each specific use case

References:

[1]. Mahmood Khalel Ibrahem ,"Modification of Diffie–Hellman Key Exchange Algorithm for Zero Knowledge Proof",2012
[2]. Preetika Joshi, Manju verma, Pushpendra R Verma , "Secure Authentication Approach Using DiffieHellman Key Exchange Algorithm for WSN" ,2015
[3]. Zhen Cheng, Yufang Huang, Jin Xu ," Algorithm for Elliptic Curve Diffie-Hellman Key Exchange Based on DNA Tile Self-assembly", 2014
[4]. Tushar Singh Chouhan, "Implementation of PRESENT Cryptographical Algorithm for the Encryption of Messages in NETFPGA 1G" ,2015
[5]. Pramod Subramanyan, Sayak Ray and Sharad Malik," Evaluating the Security of Logic Encryption Algorithms", 2015

File Transfer Using Li-Fi

Mr. Kishan Yadav, Mr.Yugesh Murugesan, Mr. Krishna Singh, Dr. Kamal Shah
(TCET-IT, Student) Mumbai, India.
kishanyadav145@gmail.com, murugesanyugesh@gmail.com, krishnas1704@gmail.com, kamal.shah@thakureducation.org

Abstract:- **With globalisation and the thirst for connectivity across society, the demand placed on wireless infrastructure and the associated resource is growing exponentially. Very soon this resource will reach saturation point, due to the finite bandwidth available in the Radio Frequency (RF) spectrum. A method of countering the impending saturation needs to be found. That method can be Visible Light Communication (VLC).**

Light Fidelity (Li-Fi) is a research field within VLC that utilises the visible light band within the electromagnetic wave spectrum. This band is 10,000 times larger than the RF band and cannot be 'leased' or saturated with users. Light waves can be modulated to carry an enormous amount of simultaneous data, at speeds faster than current consumer equipment can handle. This Dissertation describes in detail the research, construction and testing of a Li-Fi prototype using Arduino. The prototype is compact, low cost, uses accessible components and provides a solid foundation for other students to follow on with further work in this field. The prototype successfully demonstrates the principle of Visible Light Communication and shows the viability of using Python for coding, SPI for data transfer and lists suitable electronic components to process bit-wise data signals. The prototype shows that while it is possible to use addressable LED's as the transmitting element, the Dissertation concludes that they are not suitable outside of a heavily constrained environment.

Keywords— Arduino; processing; Arduino software; LED; LDR; visible light communication; file transfer.

I. INTRODUCTION

We now live in a world that is infinitely connected through a multitude of invisible, networked pathways. They stretch and travel across houses, towns, countries and continents. We are a truly global society that now has an insatiable appetite for connection, information and convenience.

This intrinsic appetite has fuel led the proliferation of connection technology from crude military beginnings at DARPA (Dennis et al. 1988), through excruciatingly slow networks of public computers and onto networks increasing in speed and computational power, now without the use of cables.

Almost exponentially the world has developed and embraced technology that now allows us to carry super-advanced, micro-computers in our pockets. The majority of these devices are connected wirelessly to internet service providers, who in turn connect to the World Wide Web. With a swipe of our finger we can find out the weather at that exact moment in Nairobi, St Petersburg and Honolulu. A supplementary or superseding method of communication needs to be developed to combat the inevitable RF band saturation. There is such a method and it utilises a relatively untapped source of waves, with an extremely large bandwidth. This method uses the visible light spectrum band, which is shown (not to scale) in Figure 1 below.

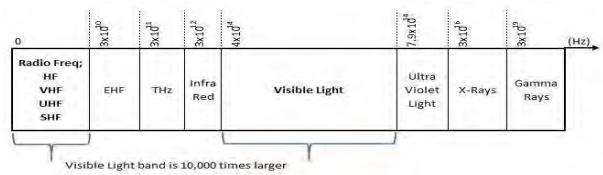


Figure 1 - Radio Frequency vs. Visible Light Frequency Bands. Courtesy of author.

Electromagnetic spectrum of RF and Visible Light are disproportionately assigned in current communication methods. Not only is there larger space to grow into, it is 'greener', the band cannot be regulated, leased or saturated, it does not create electromagnetic interference with other devices and modulation of data can occur at frequencies that the human eye cannot detect (Stefan & Haas 2014). There is even research underway by Rajagpol et al. (2012) where data streaming is being conducted at light levels such that the light appears to be off according to the human eye. This could account for daytime or purposefully dark environments where people may use their devices.

II. LITERATURE SURVEY

This literature review covers the published research relating to video or data transfer via VLC (Light Fidelity), in particular utilising Raspberry Pi or similar low-cost, consumer accessible microprocessor modules. It covers key project concepts such as any potential transmission and receiver circuit configurations, signal management, encoding and decoding of the data into a bit-stream and vice versa, and modulation techniques. VLC is a field of research that has recently taken on greater importance within our lives. Take for example, the industrial pursuit for releasing to the public, society's first truly operational 'driverless vehicle'. The foundation of this technological product is built on the foundation of VLC (coupled with radar and machine vision). The concept of transmitting data via light waves is evolving into a subject that can yield solutions across our lives, including RF band congestion and quenching the global thirst for faster and more complex data transmission.

While Dr Harald Hass demonstrated video data transmission during his TEDGlobal talk Wireless Data from Every Light Bulb (2011), the technological complexities involved in that demonstration were out of reach of anyone other than PhD level academics. Dr Haas is by far the leading worldwide exponent in the field and outside of his academic research at the University of Edinburgh, he has started his own company called pureLiFi (pureLiFi 2014). This company takes the scientific advances they test in the laboratory and turn them into

consumer orientated modules. At present they have a 10Mbps half-duplex system (Li-Flame) that works at a distance of up to 3m (pureLiFi 2014). Although this system focuses on bi-directional internet access, any development of this platform is limited to commercial partnerships with the parent company or through PhD pathways at the university. The Li-Flame system also requires additional roof modules to be mounted next to the LED lights and preclude it from being integrated into many publicly accessible applications. This is largely due to the vandalism exponent and the 'irregular' intrusion into a 'headspace' envelope. For these reasons a recessed or concealed system is mandatory for commercial acceptance.

III. METHODOLOGY

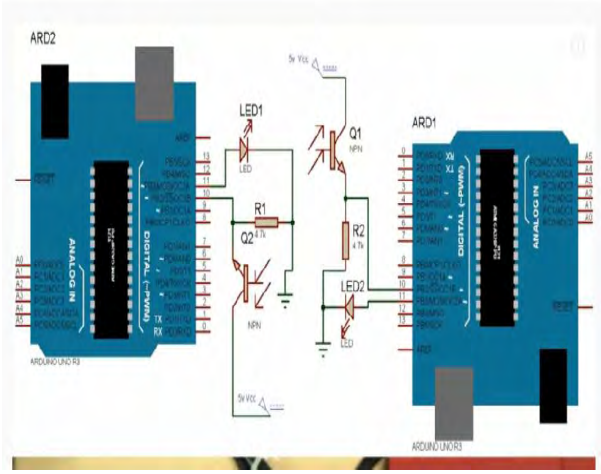


Fig.3.1 Block Diagram

This project was purely one of research and experimental work. The project was executed as a progressive development of concept following a body of research and expansion of ideas to achieve the aim. The steps generally followed a typical Project lifecycle model from the Project Management Book of Knowledge (Stackpole, 2010).

- Project Initialisation – research proposal, literature review, design information
- Project Start Up – procurement, production readiness
- Project Production Phase – build the prototype
- Project Execution – get the prototype working, test and develop
- Project Closure – write up dissertation, demobilisation

Initialisation The Project Initialisation Phase was a combination of producing a preliminary report, deeper research into prototype specifics and formalisation of the project scope and outcomes. This phase was purely research driven and the investigation carried out defined the general shape of the project. Significant information was gathered during this phase and a thorough search helped to prevent errors or miscalculations during construction. The outputs of the Initialisation Phase were;

- A mental roadmap of the Project direction, with alternative routes in reserve,
- A list of required resources, and
- Set-up of a Project Diary to track daily progress and the reasons behind the various decisions made throughout the Project.

Start-Up The Project Start-Up phase took the outputs from the Initialisation Phase and built on them to begin the procurement of Project resources, allocation of Project space, Project management systems and to mark the start of the physical work phase of the Project. This phase moved the Project from an idea, theory and research based Project to one that was tangible. Outputs of the Start-Up Phase were;

- An allocated workspace to construct the prototype,
- Focussed research to a preliminary modulation technique,
- Understood how to drive the LED's from a 3.3V Ardunio uno output signal to a LED i/p signal
- All necessary software downloaded and installed, and

Production The Production Phase marked the start of the actual building of the prototype. This was undoubtedly one of the most exciting aspects of the project and brought with it the start of the major challenge to get it working. The outputs of the Production Phase were;

- Build and develop the LED driver circuit to deliver the required voltage to modulate the LED's, and
- To completely construct all power supplies, equipment and components.

Execution This phase contained the largest amount of work, frustration, technical demand, support and investigation. The primary focus was to get it working and then to look for optimisation and improvement, time permitting. The planned sequence of work was;

1. Connected power supply and signal management circuits to drive LED's on,
2. Confirmed that photodiodes could receive the LED's being manually switched on and off,
3. Used the Arduino uno transmission module (TX) to create a simple square wave to send to the uno receiver module (RX) and achieve basic communication, displayed on a screen or other output device,
4. Increased data speed (via small steps) and maintained expected outputs,
5. Attempted to send a file across free space, maintaining the quality of the original. The outputs of the Execution Phase were;

- A Project Diary that recorded the path followed throughout the Project to provide a resource to build the Dissertation from,
- A working Prototype that could send files across free space using Visible Light Communication technology, and
- A final configuration (including software + code used) of the system.

Project Closure The Closure Phase intends to wrap up the Project and finalise all aspects ready for academic

submission. Closure includes decommissioning the prototype system, returning borrowed or loaned equipment and writing up the results. Should the Project provide suitable material, a publishable paper could be written to give information to the wider engineering community. The outputs of the Project Closure phase are;

- An academic Dissertation submitted to USQ for ENG4112,
- Decommissioning of the prototype, and
- A final parts list and configuration document to share with industry.

IV. PROJECT PROTOTYPE

Prototype Development

This chapter describes in detail the steps that were taken to arrive at a working prototype. It covers all of the technical information, principles and operation pertaining to the prototype.

Project System Diagram

The system can be broken down into two sub-systems; the Transmitter (TX) and the Receiver (RX). Each one is made up of smaller functions such as the LED electrical control circuit and board, the photodiode electrical control and board, the software in the TX and RX modules, the peripherals (keyboard, monitor, mouse etc.) and support software (e.g. MobaXterm). The Figure below illustrates how the prototype is made up from a system point of view and shows the interaction between the varying components.

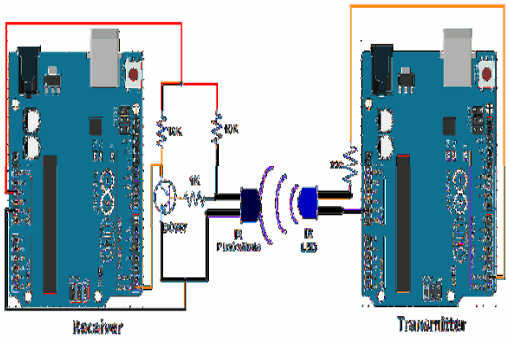


Fig.4.1 Project System Diagram

Graphical Representation

Led light is use to transfer data on basis of capacity of light to data transfer of particular led lights which is shown as follows:

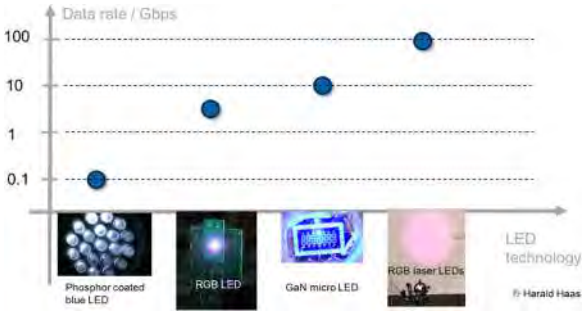


Fig.4.2 LED vs Data Rate

GUI

The system having two parts which is composed of transmitter and receiver the same thing is developed by using python code as shown follow:

Transmission Frame composed of text enter area and button to send and listbox which is used to show transmitted data to user.

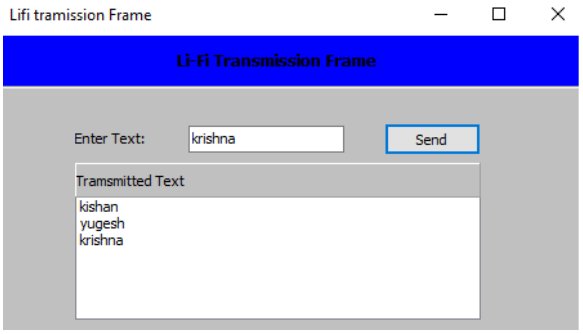


Fig.4.3.1 Transmission Frame

Receiver Frame composed of listbox which is used to show the received data.

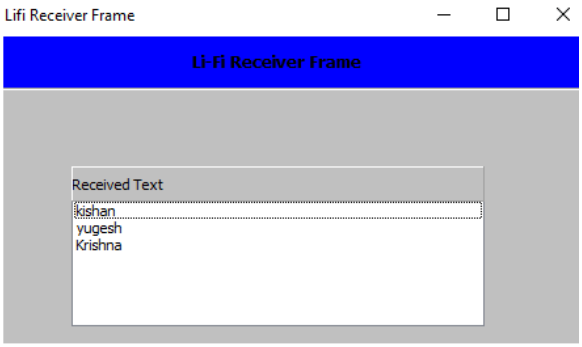


Fig.4.3.2 Receiver Frame

Port which is used for connection and baud rate is low the error rate is low.

V. FUTURE SCOPE

This Technology can be improved further and can be used to transmit the data to the mobile phones and computers that does not use radio waves for the transmission, home automation, monitoring activities and Li-Fi hotspots which can transmit the data in the order of GBPS.

VI. CONCLUSION

Therefore, by using this system, we can obtain a better speeds than Wi-Fi and normal connecting cables which are commonly used for the file transfer. The Radio band crisis problem can also be solved by employing the further existing system with this technology. As internet is not required for any of its software the bandwidth is also conserved hence reducing the network complexity. By further development of this technology can transfer the Files ranging GB's within few seconds.

VII. RESULTS

Fig.7.1 Output

1.At low baud rate the error rate is low e.g At baudrate=1200 | error <1% ,however as we increase the baudrate the error rate increases i.e At baudrate=9600 | errorr>15%.

2.With expensive LED like APA102 Dotstar with Luminous flux = 18 Lm and frequency upto 20kHz the data rate can be improved to 50kbps.

3.With Normal LED data rate upto 4-5 kbps is achievable.

4.With high performance photo receptor like HAMAMATSU S5971 Photodiode with 1Mhz frequency is acheiveable.

5.Using phototransistor (L14F1) & RS-232 protocol we are succesfully able to send text message from one PC to another PC at a rate of few kbps.

6.Though image type file is not supported in RS 232 protocol but we can still send the binary code and then at the software side process and convert it back to image.

7.To send a file of size 5kB it takes around 1-2 secs and for image of size 50kB it can take 10-12 secs.

References

1.AdaFruit Industries, 2016. About AdaFruit - Press, Limor Fried / Ladyada & more.... [ONLINE] Available at: https://www.adafruit.com/about. [Accessed 05 August 2016] AdaFruit Industries. 2016

2.Communications & Media Authority, 2013, Australian Radio Frequency Spectrum Chart 2013, ACMA, Sydney, Available at: www.acma.gov.au. [Accessed 09 October 2015].

.3. Burgess, P 2016, AdaFruit NeoPixel Uberguide. AdaFruit Learning Systems, New York. Available at: https://learn.adafruit.com/downloads/pdf/adafruit-neopixel-uberguide.pdf [Accessed 12 December 2015]. Caihong, W, Xiaohua, L & Xiuhua, J, 2013.

4. Design and Implementation of a Real-time Video Stream Analysis System Based on FFMPEG'. Fourth World Congress on Software Engineering. IEEE. Chang, Y, Liao, C, Ho, C, Wu, M, Hsieh, Y, Houng & Yang, C 2015, 'Light-emitting diodes for visible light communication', Wireless Communications and Mobile Computing Conference (IWCMC), 2015 International, pp.665-667.

5. Cheuque, C, Baeza, F, Marquez, G & Calderon, J 2015, 'Towards a Responsive Web Service for Smart Home LED Control with Raspberry Pi',34th Inernational Coference of the Chilean Computer Science Society (SCCC), 2015.

6.. Rajesh, M., and J. M. Gnanasekar. "CONGESTION CONTROL USING AODV PROTOCOL SCHEME FOR WIRELESS AD-HOC NETWORK." Advances in Computer Science and Engineering 16.1/2 (2016): 19.

7. S.V.Manikanthan and T.Padmapriya "Recent Trends In M2m Communications In 4g Networks And Evolution Towards 5g", International Journal of Pure and Applied Mathematics, ISSN NO:1314-3395, Vol-115, Issue -8, Sep 2017.

8. S.V.Manikanthan and K.Baskaran "Low Cost VLSI Design Implementation of Sorting Network for ACSFD in Wireless Sensor Network", CiiT International Journal of Programmable Device Circuits and Systems,Print: ISSN 0974 – 973X & Online: ISSN 0974 – 9624, Issue : November 2011, PDCS112011008.

9.. R. Jeevidha, V. Sowmiya , K. Kiruthiga & R. Priya, "Collabration Complexity Reducing Strategy In Cloud Computing" 124-128.

10. APA102 5050 Cool White LED w/ Integrated Driver Chip - 10 Pack [~6000K] ID: 2351 - $4.95: AdaFruit Industries, Unique & fun DIY electronics and kits. [ONLINE] Available at: https://www.adafruit.com/products/2351. [Accessed 29 August 2016].

11. GPIO Electrical Specifications, Raspberry Pi Input and Output Pin Voltage and Current Capability. [ONLINE] Available at: http://www.mosaic-industries.com/embeddedsystems/microcontroller-projects/raspberry-pi/gpio-pin-electrical-specifications.

12. Ambady, S, Bredes, M & Nquyen, C 2015, 'Visible Light Communication', BSc Thesis, Worchester Polytechnic Institute, Massachusetts,USA. Australian

13. Chung, Y, Sewaiwar, A, Vikramaditya, V & Tiwari, S 2015, 'Novel user allocation scheme for full duplex multiuser bidirectional Li-Fi network', Optics Communications, vol. 339, pp. 153-156. Clifford, P 2016.

Fingerprint Security on cloud using IOT

Niket Singh, Utsav Singh, Tanya Sinha, Neha Patwari
Information Technology, Thakur College Of Engineering and Technology, Mumbai, India
niket11097@gmail.com, singh.utsav100@gmail.com, tanyasinha017@gmail.com, neha.patwari83@gmail.com

Abstract- **The Internet of Things (IoT) is a system which consist of interrelated computing devices, mechanical and digital machines that are provided with unique identifiers (UIDs) and has the capacity to transfer data over a network without requiring human-to-human or human to-computer interaction. IoT security is the technology area concerned with safeguarding connected devices and networks in the internet of things (IoT). IoT involves adding internet connectivity to a system of interrelated computing devices, mechanical and digital machines. Each thing is provided with a unique identifier and the ability to automatically transfer important data over a network. The activities like storing the data securely and processing the data take place in the cloud rather than on the device itself, this has significant implications for IoT. Many IoT systems makes use of large numbers of sensors to collect data and then make high quality intelligent decisions. Current approaches of authentication is based on mainly password and pin based authentication. Here we have planned to implement a biometric based authentication system which would be based on fingerprint biometrics. It would be implemented using a fingerprint sensor and raspberry pi 3. The fingerprint credentials collected using the sensor would be stored in cloud such as Google, AWS, etc or even our own cloud server. As we know that major drawback of cloud is its security issue. The stored fingerprint credentials could be attacked by the intruder and can be explored further for illegal and unauthorized activities, this project stores fingerprint over the virtual cloud and provides high security to the fingerprint. The fingerprint is collected from raspberry pi 3 and stored over the virtual cloud created and the security is given using the AES or RSA algorithm. So we are focusing on providing security to the data stored over the cloud. The conventional cloud security (public cloud) has weak security as it is made feasible to large number of users.**

Keywords: Encryption, IoT, Cloud Security

I. INTRODUCTION

The reach of internet has extended itself to many applications and such usage is now clocking at intense level. To put this into perspective, a live example is Banking where lockers and confidential data are stored using the physiological characteristics. Biometric authentication is one of the popular technologies.[1] This biometric identification includes fingerprint, iris and face recognition etc. The usage of Raspberry Pi and its connectivity with cloud is now becoming a new trend. In addition to this feature we would provide security to the data on cloud using a security algorithm. As the storage on local disks is unreliable in some cases, cloud-based storage is attracting everyone to store data on cloud which can be accessible from everywhere.[3] The number of individuals storing their data on cloud is increasing day by day, so the security level needs to be upgraded because when particular data is uploaded to the cloud the information is transparent to two parties, one is cloud services and the other is cloud administrator. Everyone uses their own encryption mechanisms and keys. By this many people think that their data is secured, but no there is a chance to know the respective key of an individual.

II. BACKGROUND

Blowfish algorithm is used for encryption process for securing information. It is used for image encryption. Blowfish is a symmetric block cipher that has been utilized generally these days. On the off chance that a client needs to store his/her information in the cloud through Cloud Service Provider (CSP) the client presents their prerequisites and picks best-indicated administrations offered by the provider.[5] At the point when the information is moved to the picked CSP occurs and in future at any point if the user uploads any information in the cloud, the information will initially be encrypted using the AES algorithm and after that sent to the provider. The information is transferred on the cloud after its encryption, any demand to peruse the information will happen after it is decrypted on the client's end and after that the plain content information can be perused by the client. This incorporates a wide range of information. This encryption is straightforward to the application and can be coordinated rapidly and effectively with no progressions to the application. Since it might bargain the key additionally, the key is never put away alongside the encoded information.[1] A physical key administration server can be introduced in the client's premises so as to store the keys. This encryption assurances to secure information and keys that they stay under client's control and will never be exposed in storage or in transit. DIFFIE-HELLMAN for key exchange. It is a selected algorithm for changing cryptographic keys. It is a standout amongst the maximum activate sensible models of key trade within the area of cryptography.[4] The Diffie– Hellman key trade system grants social affairs that have no earlier studying of each different to together set up a shared mystery key over a questionable correspondences channel. This key would then have the capacity to be used to scramble subsequent trades using a symmetric key figure. The arrangement was first dispersed by Whitfield Diffie and Martin Hellman in 1976, regardless of the way that it had been freely built up several years sooner inside GCHQ, the British signs knowledge office, by James H. Ellis, Clifford Cocks, and Malcolm J. Williamson, notwithstanding, was kept

classified.[citation needed] In 2002, Hellman proposed the estimation be called Diffie– Hellman– Merkle enter the exchange affirmation of Markel's duty to the advancement of open key cryptography.

III. IMPORTANCE OF THE PROJECT

Biometric security measures exist at various points in the data pathway that makes up the IoT. From securing a device with a fingerprint sensor to utilizing a smartcard to verify your identity, these security points help create a seamless experience and allow data to flow freely and quickly between points. That data arriving untampered is important. This allows for quick and efficient connections to be made, and keeps the data used to make those connections as secure as possible during the transfer and delivery processes. As the number of connected devices continues to grow, the need for failsafe security becomes more important and will continue to stay in the forefront of security expert and developers minds. It provides real time reports. It provides interoperability with other applications, which is quite affordable, easy to install and had no maintenance cost. It is independent of any password, pin or RFID card.

IV. OBJECTIVES AND SCOPE OF THE PROJECT

1) To provide a system that gives security to biometric of user.
2) To successfully implement the security encryption technique over the cloud.
3) To develop the project such that it can be easily integrated with the new technologies.
This is an encryption system that stores the biometrics (fingerprint) over the cloud. It is more cheaper than compared to other similar system. It uses the Diffie-Hellman algorithm to encrypt the images stored which is a strong algorithm for commercial purpose.

V. SUMMARY

Here we analyzed various familiar research papers relating our undertaking as a part of the literature survey. The venture goals at growing an more secured cloud the usage of the powerful safety algorithm. This utility will be capable of help the safety and other developers' paintings efficaciously. Our literature survey became focused on the abilities of the selections we made for the middle part of the undertaking, particularly key encryption set of rules and the cloud generation that provides protection to records We have divided the challenge into 2 stages, which are in addition divided into five and a couple of sub-stages every. The trouble definitions for each section and sub-segment are as follows. 13 Phase I – Phase I will recognition on forming a base for the project. With thorough making plans, evaluation and design, we can make sure that the real implementation is smoother. Phase II – Phase II will consciousness on satisfactory tuning the task and ensuring that the supposed functions will work as we desired them to. Due to these 2 being freely available, the mission may be very low value and for that reason viable from an financial standpoint. The expected general cost is hence 300 rupees. As such, we do no longer need any outside investment. The skills required for the venture are fairly within the feasible variety. Most of the planned feature hinge around processing datasets. A cloud server is required to be able to be performed on AWS platform. The server might be kept simple and minimalist. We even have a function for users to set private records, preferences. On an operational stage, the software will want to run extensive processing duties even as it's in its getting to know phase. This utility may have many advantageous packages on society. It will permit customers to manipulate their healthcare desires in a reasonably-priced and clean manner. Since the venture deals with a touchy subject matter with potential for foremost damage, it will convey a disclaimer for mistaken utilization. The software isn't always intended to be an alternative to doctors or hospital supplied healthcare. Users must ideally best use it to supplement expert healthcare. However, the potential for misuse, whether intentional or no longer, nonetheless exists. The technique that we chose for our task is Agile. Agile lets in us to be flexible, manage converting requirements, manage the ever-increasing scope as well as get customer perspective. Agile has cognizance on client delight, which is a crucial factor of any healthcare software. Agile software program development is a technique for software program tasks below whose framework venture necessities and answers evolve via the collaborative attempt of self-organizing and go-purposeful groups and their consumer and meant customers. Users showed interest in Fingerprint safety over cloud the usage of IoT, in addition to a software which could do it for a reasonable price. Besides this, we researched pursuits of clients in wearable technologies. Technologies like KYC have already made a touch within the security marketplace.

VI. DESIGN AND IMPLEMENTATION :

Algorithm to be used: Diffie-Hellman: The Diffie-Hellman algorithm is being used to establish a shared secret that can be used for secret communications while exchanging data over a public network using the elliptic curve to generate points and get the secret key using the parameters. Step by Step Explanation:
Flowchart

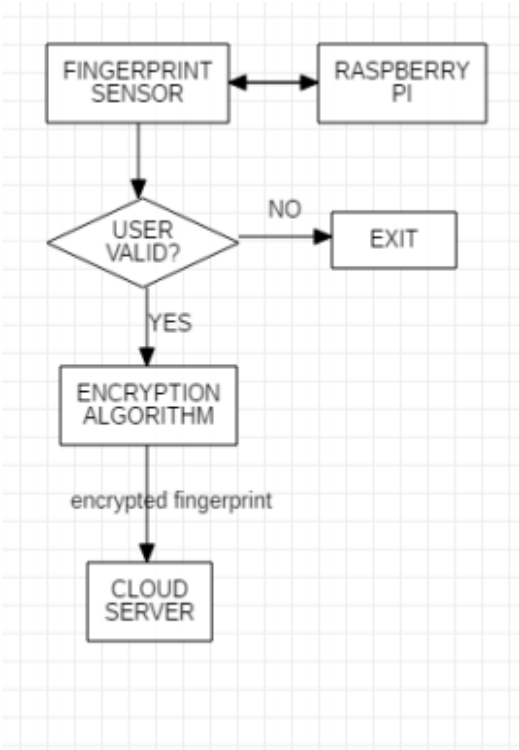


Figure: Flow Chart

Block Diagram

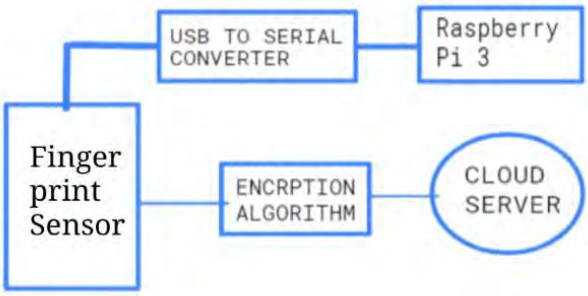


Figure: Block Diagram

Algorithm used

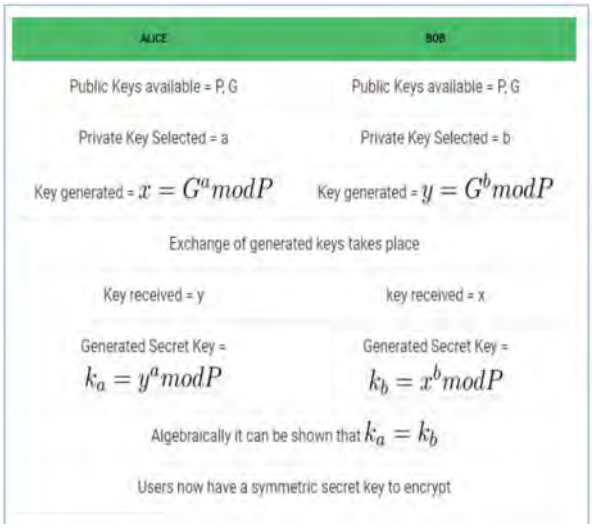


Figure: Algorithm used

Components:

Hardware:

A. Fingerprint Module

The captured fingerprint traits are sent to the Raspberry Pi has enhancement and fingerprint verification for capturing fingerprints. Figure 3 Block diagram of fingerprint based digital transaction system. The main stages of this algorithm include normalization, ridge orientation estimation, ridge frequency estimation and filtering. The first step in this approach involves the normalization of the fingerprint image so that it has a pre-specified mean and variance. An orientation image is then calculated, which is a matrix of direction vectors representing the ridge orientation at each location in the image.[2] The next step in the image enhancement process is the estimation of the ridge frequency image. The frequency image defines the local frequency of the ridges contained in the fingerprint. The next step in the enhancement process is to construct the final filtered image using the pixel values from the prefiltered images.[1] Lastly, local adaptive 50 thresholding is applied to the directionally filtered image, which produces the final enhanced binary image. After a fingerprint image has been enhanced, the next step is to extract the minutiae from the enhanced image. Following the extraction of minutiae, a final image post processing stage is performed to eliminate false minutiae.

B. Raspberry Pi

Raspberry Pi has authenticated the user (by using right thumb) by using fingerprint matching algorithm and displays the user details on the screen. Next to perform the digital transactions, Raspberry Pi has generated the random finger using novel infinity noise algorithm. Then Raspberry Pi verify the injected random finger and perform the operations like balance enquiry, money transfer, money withdraw and deposit the money. If the finger hasn't match the system asks right thumb for authentication and system start the procedure from beginning. Now the security of our money is literally in our hands or rather on our fingertips.[3] Enter the password to open digital bank account with the help of a keypad. Immediately the bank account will be opened. After the work has been completed if key is pressed

again with help of touch pad the account will be closed again or it close automatically after 3 minutes if the user not processing any data. If an unauthorized person tries to scan his fingerprint image then an indication will be given by a buzzer which is interfaced to the controller and also if wrong password is entered by the user again indication will be given by the buzzer.

C. USB to Serial convertor:

A USB to serial adapter, also referred to as a USB serial converter or RS232 adapter is a small electronic device which can convert a USB signal to serial RS232 data signals. RS232 is the type of signal which is in many older PCs and is referred to as a serial COM port.

A USB to serial adapter typically converts between USB and either RS232, RS485, RS422 or TCP signals, however some USB to serial adapters have other special conversion features such as custom baud rates, high-speed or other. Even the USB to serial adapter RS232 standard is an older communication protocol it is still used by many modern serial RS232 devices in both business and consumer markets and is also often used for personal and office serial devices. [3] More recently most new computers do not have a built-in COM port so a USB serial adapter is often used for connecting many types of serial devices to a computer. A standard USB to serial adapter is a very useful device for connecting equipment such as printers, scanners, scales and GPS devices, but also most business and consumer equipment can be connected to a computer by using an industrial grade adapter.

Software:

A. 000Webhost:

000webhost is one of the very few web hosts which give you the ability to host your website while paying nothing. This may sound like a hoax or a fraud but 000webhost has been in the paid hosting business for around 4 years and they are providing free hosting since 2 years. 000webhost is a great place to start experimentations or hosting relatively unimportant websites. The free hosting which is provided by 000webhost does have some drawbacks which can be instantaneously removed by upgrading your account. 000webhost or $0.00 webhost as they liked to be called have a very reliable performance history and they also guarantee 99% uptime.

B. Python IDE:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python is a MUST for students and working professionals to become a great Software Engineer specially when they are working in Web Development Domain.

Advantages of learning Python:

- Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- Python is a Beginner's Language − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

VII. RESULTS AND DISCUSSION

a. Outcomes:

1) The project made is compatible with any environment with little atomization.
2) The project will carry out process as programmed for the security purposes.

b. Discussion of the results:

Finally this project will store the biometric (fingerprint) of the user and provide it security using the key encryption algorithm. The fingerprint is captured using the fingerprint sensor which is connected to raspberry pi 3B model. This data will be encrypted using Diffe-Hellman algorithm to encrypt the fingerprint. This system uses the
Linux as the environment for the raspberry pi and python for programming. The cloud server is made using the programming language and then the security algorithm is applied to it. This system makes a secure environment for the storage over the cloud which in current scenario is not given attention.

VIII. CONCLUSION

The proposed Project has high quality value in terms of research point. The presented model is low-cost IoT

based Biometric verification system with the use of Raspberry pi. The proposed system is efficient / secure and can be used for security and access control mechanisms and can be implied in any areas where authentication is must essential. However, we will able to achieve greater and Extra versatility, which was probable the motive that the algorithm changed into no longer that accurate inside the first region. Fingerprint security using IoT provides a solution for the security concerns of cloud services and tries to overcome that problem. IoT system faces more security concerns than normal IT system. IoT system contains new communication technologies and resource constrained devices. There is also different attack methods involved in IoT system than that of normal information system. Security on IoT system can be deployed by presenting a security framework according to the security need of IoT system. Security framework helps in understanding and minimizing security threats and issues of IoT system. Certified authorities can provide trust system for mutual authentication between multiple end points. By using digital certificates, keys and security credentials, authenticated and encrypted communication can be made possible. Our future work will address a special issue of back up connection in IoT system along with implementation details.

References:

[1] He Li, Kaoru Ota , Mianxiong Dong, “Learning IoT in edge: Deep learning for the Internet of Things with Edge Computing”,2018.

[2] Leith Abed, Nathan Clarke, Bogdan Ghita1, Abdulrahman Alruban “Securing Cloud Storage by Transparent Biometric Cryptography”,2019.

[3] Shipashree P.S, Abhishek Kumar Tiwari, Ashutosh Prakash, Surabh Kumar Singh “Cloud Based Secured Locker”,2019.

[4] Hisham Al-Assam, Walid Hasan, Sherali Zeadally “Automated Biometric Authentication with Cloud Computing”,2019.

[5] Rishab Garg, Anisha Arora, Saurabh Singh, Shipra Sarawat “Biometric Authentication using soft Biometric Traits” ,2018.

[6] Implementing Digital Signature with RSA Encryption to decorate statistics protection in cloud,2010.Uma Somani, Kanika Lakhani, Manish Mudra.

[7] Fingerprint Authentication using Raspberry Pi based on IOT. Radha N, Vijaysanthi. R, Sindhujaa V, Jaya Shree M.

[8] Anil K. Jain, Karthik Nandakumar, and Abhishek Nagar, Review Article: Biometric Template Security, Journal on Advances in Signal Processing Volume 2008, Article ID 5796.

Tiny Encryption Algorithm for IOT Data Security

Mr. Ashutosh Pandey, Ms. Sneha Thakur, Mr. Dushyant Trivedi
Thakur College Of Engineering. & Technology
Mumbai, India
pandeyashutosh2498@gmail.com, snehathakur0578@gmail.com, dushyant.trivedi99@gmail.com

Abstract— **In wireless technology we have created an exponential rise in the number of connected devices leading to the internet of things (IoT) revolution. Large amounts of data are captured, processed and transmitted through the network by these embedded devices (An embedded device is an object that contains a special-purpose computing system. The system, connected to the internet which is completely enclosed by the object, may or may not be able to connect to the Internet).**
In Iot networks Security of the transmitted data is a major concern. Encryption algorithm such as:DES/3DES or TripleDES, TEA(Tiny Encryption Algorithm), Blowfish, AES, Two fish, IDEA, MD5, SHA 1, HMAC, etc. The lower memory utilization and ease of implementation on both hardware and software scales in TEA is the most attractive among all. The usage of the same key through all rounds of encryption, produces a reduced security evident from the avalanche effect of the algorithm is the major issue in TEA. Also,at the time of encryption and decryption the text is high, leading to lower efficiency in IoT networks with embedded devices. This project uses the symmetric encryption algorithm (TEA) which is proposed in order to minimize the memory footprint, processing time and maximize the speed, increasing the performance of the system. The project is targeted for the embedded system or Smart mobile phone which concerns more over speed & space. The basic operations from mixed algebraic groups and a large number of rounds to attain security with severity the plaintext is encrypted and decrypted. It has sixty- four (64) Feistel rounds are used in the TEA encryption as well as decryption process. The outcomes is to design a possible implementation on embedded devices through which cloud services will store the data passed by the devices. These projects can be possibly taken further from embedded devices to Smartphone devices.

Keywords—IoT; Security; Encryption; Wireless Sensor Network WSN; Khazad

I.INTRODUCTION

The Internet of Things (IoT) is turning out to be an emerging discussion in the field of research and practical implementation in the recent years. IoT is a model that includes ordinary entities with the capability to sense and communicate with fellow devices using Internet [1]. As the broadband Internet is now generally accessible and its cost of connectivity is also reduced, more gadgets and sensors are getting connected to it [2]. Such conditions are providing suitable ground for the growth of IoT. There is great deal of complexities around the IoT, since we wish to approach every object from anywhere in the world [3]. The sophisticated chips and sensors are embedded in the physical things that surround us, each transmitting valuable data. The process of sharing such large amount of data begins with the devices themselves which must securely communicate with the IoT platform. This platform integrates the data from many devices and apply analytics to share the most valuable data with the applications. The IoT is taking the conventional internet, sensor network and mobile network to another level as every thing will be connected to the internet. A matter of concern that must be kept under consideration is to ensure the issues related to confidentiality, data integrity and authenticity that will emerge on account of security and privacy [4].

A. Applications of IoT:

With the passage of time, more and more devices are getting connected to the Internet. The houses are soon to be equipped with smart locks [5], the personal computer, laptops, tablets, smart phones, smart TVs, video game consoles even the refrigerators and air conditioners have the capability to communicate over Internet. This trend is extending outwards and it is estimated that by the year 2020 there will be over 50 billion objects connected to the Internet [6]. This estimates that for each person on earth there will be 6.6 objects online. The earth will be blanketed with millions of sensors gathering information from physical objects and will upload it to the Internet. It is suggested that application of IoT is yet in the early stage but is beginning to evolve rapidly [7], [8]. An overview of IoT in building automation system is given in [9]. It is suggested in [10] that various industries have a growing interest towards use of IoT. Various applications of IoT in health care industries are

discussed in [11], [12] and the improvement opportunities in health care brought in by IoT will be enormous [13].

It has been predicted that IoT will contribute in the making the mining production safer [14] and the forecasting of disaster will be made possible. It is expected that IoT will transform the automobile services and transportation systems [15]. As more physical objects will be equipped with sensors and RFID tags transportation companies will be able to track and monitor the object movement from origin to destination [16], thus IoT shows promising behavior in the logistics industry as well.

With so many applications eying to adapt the technology with the intentions to contribute in the growth of economy health care facility, transportation and a better life style for the public, IoT must offer adequate security to their data to encourage the adaptation process.

B. Security Challenges in IoT:

To adopt the IoT technology it is necessary to build the confidence among the users about its security and privacy that it will not cause any serious threat to their data integrity, confidentiality and authority. Intrinsically IoT is vulnerable to various types of security threats, if necessary security measures are not taken there will be a threat of information leakage or could prove a damage to economy [17], [18]. Such threats may be considered as one of the major hindrance in IoT [19], [20].

IoT is extremely open to attacks [21], [22], for the reasons that there is a fair chance of physical attack on its components as they remain unsupervised for long time. Secondly, due to the wireless communication medium, the eavesdropping is extremely simple. Lastly the constituents of IoT bear low com- petency in terms of energy with which they are operated and also in terms of computational capability. The implementation of conventional computationally expensive security algorithms will result in the hindrance on the performance of the energy constrained devices.

It is predicted that substantial amount of data is expected to be generated while IoT is used for monitoring purposes and it is vital to preserve unification of data [23]. Precisely, data integrity and authentication are the matters of concern.

From a high level perspective, IoT is composed of three components namely, Hardware, Middleware and Presentation [1]. Hardware consists of sensors and actuators, the Middle- ware provides storage and computing tools and the presenta- tion provides the interpretation tools accessible on different platforms. It is not feasible to process the data collected from billions of sensors, context-aware Middleware solutions are proposed to help a sensor decide the most important data for processing [24]. Inherently the architecture of IoT does not offer sufficient margin to accomplish the necessary actions involved in the process of authentication and data integrity. The devices in the IoT such as RFID are questionable to achieve the fundamental requirements of authentication process that in- cludes constant communication with the servers and exchange messages with nodes.

In secure systems the confidentiality of the data is main- tained and it is made sure that during the process of message exchange the data retains its originality and no alteration is unseen by the system. The IoT is composed of many small devices such as RFIDs which remain unattended for extended times, it is easier for the adversary to access the data stored in the memory [25]. To provide the immunity against Sybil attacks in RFID tags, received signal strength indication (RSSI) based methodologies are used in [26], [27], [28] and [29].

Many solutions have been proposed for the wireless sensor networks which consider the sensor as a part of Internet connected via nodes [30]. However, in IoT the sensor nodes themselves are considered as the Internet nodes making the authentication process even more significant. The integrity of the data also becomes vital and requires special attention towards retaining its reliability.

C. Motivation And Organization of Paper:

Recently a study by HP reveals that 70% of the devices in IoT are vulnerable to attacks [31]. An attack can be performed by sensing the communication between two nodes which is known as a man-in-the-middle attack. No reliable solution has been proposed to cater such attacks. Encryption however could lead to minimize the amount of damage done to the data integrity. To assure data unification while it is stored on the middle ware and also during the transmission it is necessary to have a security mechanism. Various cryptographic algorithms have been developed that addresses the said matter, but their utilization in IoT is questionable as the hardware we deal in the IoT are not suitable for the implementation of computationally expensive encryption algorithms. A trade-off must be done to fulfil the requirement of security with low computational cost.

In this paper, we proposed a lightweight cryptographic algorithm for IoT named as Secure IoT (SIT). The proposed algorithm is designed for IoT to deal with the security and resource utilization challenges mentioned

in section I-B. The rest of the paper is organized as follows, in section II, a short literature review is provided for the past and contemporary lightweight cryptographic algorithms, in section III, the detail architecture and functioning of the proposed algorithm is pre- sented. Evaluation of SIT and experimental setup is discussed in section V. Conclusion of the paper is presented in section VII.

II.CRYPTOGRAPHIC ALGORITHMS FOR IOT:

The need for the lightweight cryptography have been widely discussed [32], [33], [34], also the shortcomings of the IoT in terms of constrained devices are highlighted. There in fact exist some lightweight cryptography algorithms that does not always exploit security-efficiency trade-offs. Amongst the block cipher, stream cipher and hash functions, the block ciphers have shown considerably better performances.

A new block cipher named mCrypton is proposed [35]. The cipher comes with the options of 64 bits, 96 bits and 128 bits key size. The architecture of this algorithm is followed by Crypton [36] however functions of each component is simplified to enhance its performance for the constrained hardware. In [37] the successor of Hummingbird-1 [38] is proposed as Hummingbird-2(HB-2). With 128 bits of key and a 64 bit initialization vector Hummingbird-2 is tested to stay unaffected by all of the previously known attacks. However the cryptanalysis of HB-2 [39] highlights the weaknesses of the algorithm and that the initial key can be recovered. [40] studied different legacy encryption algorithms including RC4, IDEA and RC5 and measured their energy consumption. They computed the computational cost of the RC4 , IDEA and RC5 ciphers on different platforms. However, various existing algorithms were omitted during the study.

TEA , Skipjack and RC5 algorithms have been implemented on Mica2 hardware platform . To measure the energy consumption and memory utilization of the ciphers Mica2 was configured in single mote. Several block ciphers in- cluding AES , XXTEA , Skipjack and RC5 have been implemented, the energy consumption and execution time is measured. The results show that in the AES algorithm the size of the key has great impact on the phases of encryption, Decryption and key setup i-e the longer key size results in extended execution process. RC5 offers diversified parameters i-e size of the key, number of rounds and word size can be altered. Authors have performed variety of combinations to find out that it took longer time to execute if the word size is increased. Since key setup phase is not involved in XXTEA and Skipjack, they drew less energy but their security strength is not as much as AES and RC5. proposed lightweight block cipher Simon and Speck to show optimal results in hardware and software respectively. Both ciphers offer a range of key size and width, but atleast 22 numbers of round require to perform sufficient encryption. Although the Simon is based on low multiplication complexity but the total number of required mathematical operation is quite high.

III.PROPOSEd SYSTEM

The architecture of the proposed algorithm provides a simple structure suitable for implementing in IoT environment. Some well known block cipher including AES (Rijndael) , 3-Way , Grasshopper , PRESENT , SAFER, SHARK , and Square use SubstitutionPermutation (SP) network. Several alternating rounds of substitution and transposition satisfies the Shannon's confusion and diffusion properties that ensues that the cipher text is changed in a pseudo random manner. Other popular ciphers including SF, Blowfish, Camelia and DES , use the feistel architecture. One of the major advantage of using feistel architecture is that the encryption and decryption operations are almost same. The proposed algorithm is a hybrid approach based on feistel and SP networks. Thus making use of the properties of both approaches to develop a lightweight al- gorithm that presents substantial security in IoT environment while keeping the computational complexity at moderate level.

Another vital process in symmetric key algorithms is the generation of key. The key generation process involves complex mathematical operations. In WSN environment these operations can be performed wholly on decoder ,on the contrary in IoT the node themselves happens to serve as the Internet node, therefore, computations involved in the process of key generation must also be reduced to the extent that it ensures necessary security. In the sub-sections the process of key expansion and encryption are discussed in detail. Some notations used in the explanation are shown in Table I

TABLE I: Notations

Notation	Function
⊥	XOR
J	XNOR
++, ‖	Concatenation

KEY EXPANSION

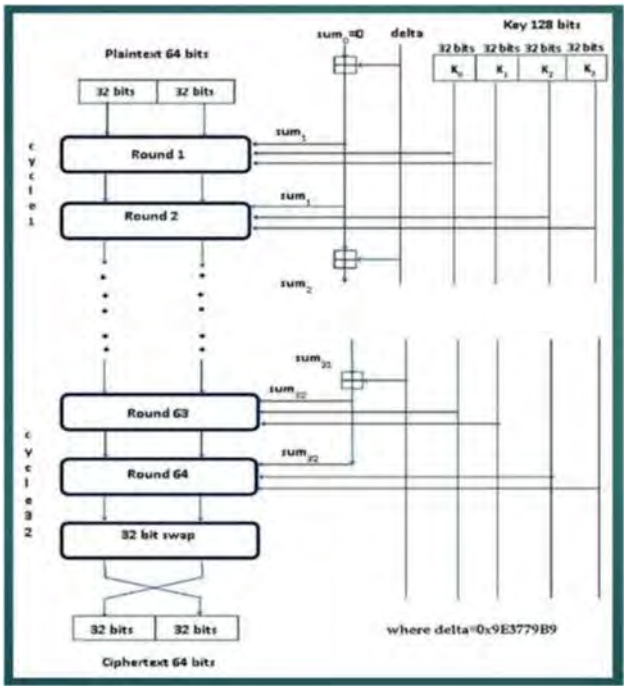


Fig. 1: Key Expansion

The most fundamental component in the processes of encryption and decryption is the key. It is this key on which entire security of the data is dependent, should this key be known to an attacker, the secrecy of the data is lost. Therefore necessary measures must be taken into account to make the revelation of the key as difficult as possible. The feistel based encryption algorithms are composed of several rounds, each round requiring a separate key. The encryption/decryption of the proposed algorithm is composed of five rounds, therefore, we require five unique keys for the said purpose. To do so, we introduce a key expansion block which is described in this section.

To maintain the security against exhaustive search attack the length of the true key kt must be large so that it becomes beyond the capability of the enemy to perform $2kt^{-1}$ encryp- tions for key searching attacks. The proposed algorithm is a 64- bit block cipher, which means it requires 64-bit key to encrypt 64-bits of data. A cipher key (Kc) of 64-bits is taken as an input from the user. This key shall serve as the input to the key expansion block. The block upon performing substantial operations to create confusion and diffusion in the input key will generate five unique keys. These keys shall be used in the encryption/decryption process and are strong enough to remain indistinct during attack. The architecture of the key expansion block is shown in Fig. 1. The block uses an f -function which is influenced by tweaked Khazad block cipher . Khazad is not a feistel cipher and it follows wide trial strategy. The wide trial strategy is composed of several linear and non-linear transformations that ensures the dependency of output bits on input bits in a complex manner. Detailed explanation of the components of key expansion are discussed below:

In the first step the 64-bit cipher key (Kc) is divided into the segments of 4-bits.

The f -function operates on 16-bits data. Therefore four f -function blocks are used. These 16-bits for each f - function are obtained after performing an initial substitu- tion of segments of cipher key (Kc) as shown in equation

IV.ENCRYPTION & DECRYPTION

A. TEA Encryption Function

```
void encrypt(unsigned long k[], unsigned long text[])
{
unsigned long y = text[0], z = text[1];

unsigned long delta = 0x9e3779b9, sum = 0; int n;

for (n= 0; n < 32; n++) {

sum += delta;

y += ((z << 4) + k[0]) ^ (z+sum) ^ ((z >> 5) + k[1]);

z += ((y << 4) + k[2]) ^ (y+sum) ^ ((y >> 5) + k[3]);

}
```

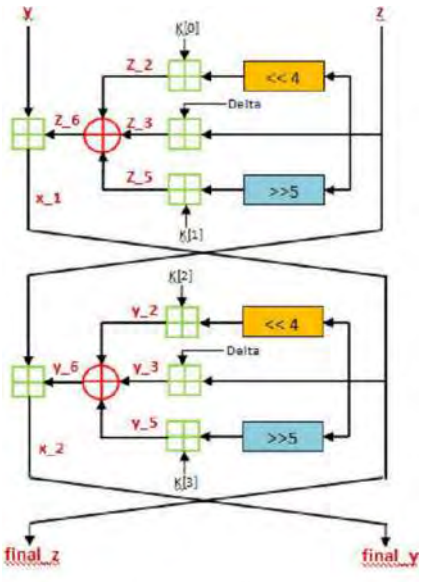


Fig. 1: Encryption Function

B. TEA decryption function

```
void decrypt(unsigned long k[], unsigned long text[])
{
unsigned long y = text[0], z = text[1];
unsigned long delta = 0x9e3779b9, sum = delta << 5;
int n;
for (n= 0; n < 32; n++) {
z -= ((y << 4) + k[2]) ^ (y + sum) ^ ((y >> 5) + k[3]);
y -= ((z << 4) + k[0]) ^ (z + sum) ^ ((z >> 5) + k[1]);
sum -= delta;  }
                text[0] = y; text[1] = z;  }
```

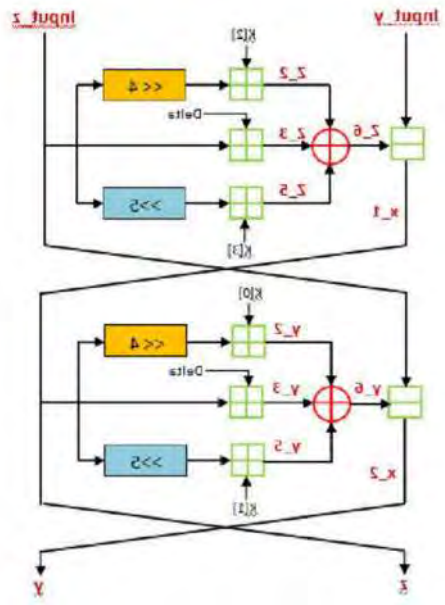


Fig. 1: Decryption Function

V.CONCLUSION

In the near future Internet of Things will be an essential element of our daily lives. Numerous energy constrained devices and sensors will continuously be communicating with each other the security of which must not be compromised. The implementation gives desired results which makes the algorithm accepted in IoT applications. In the near future we are interested in the detail performance evaluation and cryptanalysis of this algorithm on different hardware and software platforms for possible attacks.

Reference

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," Future Generation Computer Systems, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] R. Want and S. Dustdar, "Activating the internet of things [guest editors' introduction]," Computer, vol. 48, no. 9, pp. 16–20, 2015. [3] J. Romero-Mariona, R. Hallman, M. Kline, J. San Miguel, M. Major, and L. Kerr, "Security in the industrial internet of things," 2016.

[4] H. Suo, J. Wan, C. Zou, and J. Liu, "Security in the internet of things: a review," in Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on, vol. 3. IEEE, 2012, pp. 648–651.

[5] G. Ho, D. Leung, P. Mishra, A. Hosseini, D. Song, and D. Wagner, "Smart locks: Lessons for securing commodity internet of things devices," in Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security. ACM, 2016, pp. 461–472.

[6] D. Airehrour, J. Gutierrez, and S. K. Ray, "Secure routing for internet of things: A survey," Journal of Network and Computer Applications, vol. 66, pp. 198–213, 2016.

[7] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," Ad Hoc Networks, vol. 10, no. 7, pp. 1497–1516, 2012. [8] L. Da Xu, "Enterprise systems: state-of-the-art and future trends," IEEE Transactions on Industrial Informatics, vol. 7, no. 4, pp. 630–640, 2011.

[9] P. Zhao, T. Peffer, R. Narayanamurthy, G. Fierro, P. Raftery, S. Kaam, and J. Kim, "Getting into the zone: how the internet of things can improve energy efficiency and demand response in a commercial building," 2016.

[10] Y. Li, M. Hou, H. Liu, and Y. Liu, "Towards a theoretical framework of strategic decision, supporting capability and information sharing

[11] Z. Pang, Q. Chen, J. Tian, L. Zheng, and E. Dubrova, "Ecosystem analysis in the design of open platform-based in-home healthcare terminals towards the internet-of-things," in Advanced Communication Technology (ICACT), 2013 15th International Conference on. IEEE, 2013, pp. 529–534. [12] S. Misra, M. Maheswaran, and S. Hashmi, "Security challenges and approaches in internet of things," 2016.

[13] M. C. Domingo, "An overview of the internet of things for people with disabilities," Journal of Network and Computer Applications, vol. 35, no. 2, pp. 584–596, 2012.

[14] W. Qiuping, Z. Shunbing, and D. Chunquan, "Study on key technologies of internet of things perceiving mine," Procedia Engineering, vol. 26, pp. 2326–2333, 2011.

[15] H. Zhou, B. Liu, and D. Wang, "Design and research of urban intelligent transportation system based on the internet of things," in Internet of Things. Springer, 2012, pp. 572–580.

[16] B. Karakostas, "A dns architecture for the internet of things: A case study in transport logistics," Procedia Computer Science, vol. 19, pp. 594–601, 2013.

[17] H. J. Ban, J. Choi, and N. Kang, "Fine-grained support of security services for resource constrained internet of things," International Journal of Distributed Sensor Networks, vol. 2016, 2016.

[18] S. Khan, M. Ebrahim, and K. A. Khan, "Performance evaluation of secure force symmetric key algorithm," 2015. [19] P. L. L. P. Pan Wang, Professor Sohail Chaudhry, S. Li, T. Tryfonas, and H. Li, "The internet of things: a security point of view," Internet Research, vol. 26, no. 2, pp. 337–359, 2016. [20] M. Ebrahim, S.

Khan, and U. Khalid, “Security risk analysis in peer 2 peer system; an approach towards surmounting security challenges,” arXiv preprint arXiv:1404.5123, 2014. [21] M. A. Simplicio Jr, M. V. Silva, R. C. Alves, and T. K. Shibata, “Lightweight and escrow-less authenticated key agreement for the internet of things,” Computer Communications, 2016.

[22] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” Computer networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[23] F. Xie and H. Chen, “An efficient and robust data integrity verification algorithm based on context sensitive,” way, vol. 10, no. 4, 2016.

[24] S. Wang, Z. Zhang, Z. Ye, X. Wang, X. Lin, and S. Chen, “Application of environmental internet of things on water quality management of urban scenic river,” International Journal of Sustainable Development & World Ecology, vol. 20, no. 3, pp. 216–222, 2013.

[25] T. Karygiannis, B. Eydt, G. Barber, L. Bunn, and T. Phillips, “Guidelines for securing radio frequency identification (rfid) systems,” NIST Special publication, vol. 80, pp. 1–154, 2007. [26] J. Wang, G. Yang, Y. Sun, and S. Chen, “Sybil attack detection based on rssi for wireless sensor network,” in 2007 International Conference on Wireless Communications, Networking and Mobile Computing. IEEE, 2007, pp. 2684–2687.

[27] S. Lv, X. Wang, X. Zhao, and X. Zhou, “Detecting the sybil attack cooperatively in wireless sensor networks,” in Computational Intelligence and Security, 2008. CIS’08. International Conference on, vol. 1. IEEE, 2008, pp. 442–446.

[28] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, “Detecting and localizing identity-based attacks in wireless and sensor networks,” IEEE Transactions on Vehicular Technology, vol. 59, no. 5, pp. 2418–2434, 2010.

[29] S. Chen, G. Yang, and S. Chen, “A security routing mechanism against sybil attack for wireless sensor networks,” in Communications and Mobile Computing (CMC), 2010 International Conference on, vol. 1. IEEE, 2010, pp. 142–146.

[30] L. Eschenauer and V. D. Gligor, “A key-management scheme for distributed sensor networks,” in Proceedings of the 9th ACM conference on Computer and communications security. ACM, 2002, pp. 41–47.

[31] S. A. Kumar, T. Vealey, and H. Srivastava, “Security in internet of things: Challenges, solutions and future directions,” in 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 2016, pp. 5772–5781.

[32] M. Katagi and S. Moriai, “Lightweight cryptography for the internet of things,” Sony Corporation, pp. 7–10, 2008. [33] M. Ebrahim, S. Khan, and S. S. U. H. Mohani, “Peer-to-peer network simulators: an analytical review,” arXiv preprint arXiv:1405.0400, 2014.

[34] M. Ebrahim, S. Khan, and U. B. Khalid, “Symmetric algorithm survey: A comparative analysis,” International Journal of Computer Applications (0975 – 8887), vol. 61, no. 20, 2014. [35] C. H. Lim and T. Korkishko, “mcrypton–a lightweight block cipher for security of low-cost rfid tags and sensors,” in Information Security Applications. Springer, 2005, pp. 243–258. [36] C. H. Lim, “Crypton: A new 128-bit block cipher,” NIsT AEs Proposal, 1998.

[37] D. Engels, M.-J. O. Saarinen, P. Schweitzer, and E. M. Smith, “The hummingbird-2 lightweight authenticated encryption algorithm,” in RFID. Security and Privacy. Springer, 2011, pp. 19–31.

[38] D. Engels, X. Fan, G. Gong, H. Hu, and E. M. Smith, “Ultralightweight cryptography for low-cost rfid tags: Hummingbird algorithm and protocol,” Centre for Applied Cryptographic Research (CACR) Technical Reports, vol. 29, 2009. [39] K. Zhang, L. Ding, and J. Guan, “Cryptanalysis of hummingbird-2.” IACR Cryptology ePrint Archive, vol. 2012, p. 207, 2012.

[40] P. Ganesan, R. Venugopalan, P. Peddabachagari, A. Dean, F. Mueller, and M. Sichitiu, “Analyzing and modeling encryption overhead for sensor network nodes,” in Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications. ACM, 2003, pp. 151–159.

Smart Dustbin and Garbage Monitoring System using Internet of Things (IOT)

Mr. Nikhil Singh, Mr. Ritesh Singh, Mr. Shivam Singh
Information Technology, Thakur College of Engineering and Technology, Mumbai 400101
contactnikhil1801@gmail.com, ritesh.2411.singh@gmail.com, shivam998750@gmail.com

Abstract – **Internet of Things has been more deeply embedded in our day-to-day lives at home and work. In the recent decades, Urbanization has increased tremendously. At the same time there is an increase in quantity of Garbage. Waste management has been one of the most important issues to be considered. So, it is necessarily having control on overflowing garbage as well as introducing or building a system which will overcome issue of overflowing garbage using Internet of Things (IOT). One of the major problems faced by most of the countries is the issue of increasing garbage which gets difficult to monitor giving rise to overflowing Dustbins; this is one of the motivations for this research, to deploy computing techniques in creating a barrier to wastage in order to help the environment. Smart Dustbin and Garbage Monitoring system is an innovative system which will inform the users about the level of garbage present in the Dustbin and will prevent it from overflowing. It also automatically detects the user and opens the lid. To demonstrate this the system makes use of Dustbin, where there are two ultrasonic sensors where the first ultrasonic sensor is placed at the top of Dustbin Facing Downward to detect the Garbage level and compare it with the Dustbin's Depth and other one is placed at front of dustbin and it detects the user and automatically opens the lid. The system makes use of Arduino Uno, Servomotor for automation, Bluetooth Module for sending data about the level of garbage present in dustbin to cell phone and a buzzer. The garbage level is highlighted as distance from the bottom in centimeters to show the level of garbage present in the container with the help of a display to the user. The buzzer starts ringing when the set limit of the garbage is crossed. Thus, this system helps to prevent the overflowing bins by informing about the garbage levels of the Dustbin.**

Keywords: Bluetooth Module, Arduino Uno, Garbage Monitoring.

I .INTRODUCTION

India is getting buried beneath hills of rubbish as the nation has been producing more than 1.50 lakh metric ton (MT) of strong squander each day. More awful - around
90 per cent (1,35,000 MT per day) of the overall sum is collected waste[1]. Nearly 15,000 MT of trash stay uncovered each day, coming about in nearly 55 lakh MT of strong squander arranged in open ranges each year, which leads to "serious" contamination level. Of the whole collected squander, as it were 20 per cent (27,000 MT per day) is handled and the remaining 80 per cent (1,08,000 MT per day) is dumped in landfill destination Dustbins (or Trash cans, Waste Cans) are little plastic (or metal) holders that are utilized to store junk (or squander) on a brief premise. They are regularly utilized in homes, workplaces, roads, parks etc. to gather the waste. In some places, littering could be a genuine offense and thus Open Waste Containers are the as it were way to arrange little waste. Usually, it may be a common hone to utilize isolated containers for collecting damp or dry, recyclable or non-recyclable waste. In this extend, we have outlined a straightforward framework called Shrewd Dustbin and rubbish monitoring framework utilizing Arduino, Ultrasonic Sensor, Bluetooth module and Servo Engine, where the top of the dustbin will naturally open itself upon discovery of human hand and screens the level of rubbish within the dustbin and in case the dustbin is full buzzer begins buzzing and notice is appeared in versatile.[2]

II . PROBLEM DEFINITION

This extend combats the issue of flooding strong squander containers which contaminate the surroundings. The level of trash show in any container is decided by the ultrasonic distance measuring sensor. When the trash level in any waste container surpasses a pre-defined level, then the microcontroller send a caution message to the e-monitoring station, and, the workstation at that point allots the closest waste collecting truck to gather the trash from such bins, which have sent an alarm message. It advises when the holder is at full capacity and when it must be purged, hence permitting the sanitation masters to work more efficiently and cut pointless costs.[4]

III. SYSTEM ARCHITECTURE

The IOT Garbage Monitoring System framework may be a exceptionally innovative system which is able offer assistance to keep the cities clean. This system screens the rubbish containers and educates around the level of rubbish collected within the rubbish containers through a web page. For this the framework employments ultrasonic sensors placed over the canisters to identify the trash level and compare it with the trash canisters profundity. The framework makes utilize of Arduino family microcontroller, LCD screen, Wi-Fi modem for sending information and a buzzer. The framework is powered by a 12V transformer. The LCD screen is utilized to display the status of the level of garbage collected in the bins. [1]

Whereas a web page is built to appear the status to the usermonitoring it. The internet page gives a graphical see of the garbage canisters and highlights the rubbish collected in color in arrange to appear the level of rubbish collected. The LCD screen appears the status of the trash level. The framework puts on the buzzer when the level of waste collected crosses the set restrain. Hence this framework makes a difference to keep the. City clean by educating approximately the rubbish levels of the canisters by giving

graphical picture of the canisters through a web page. The ESP8266 Wi-Fi Module may be a self-contained SOC with coordinates TCP/IP convention stack that can allow any microcontroller get to your Wi-Fi organize. The ESP8266 is able of either facilitating an application or offloading all Wi-Fi organizing capacities from another application processor. Each Bluetooth module comes preprogramed with an AT command set firmware. The Bluetooth module is an amazingly fetched compelling board with a gigantic, and ever developing localities[3].

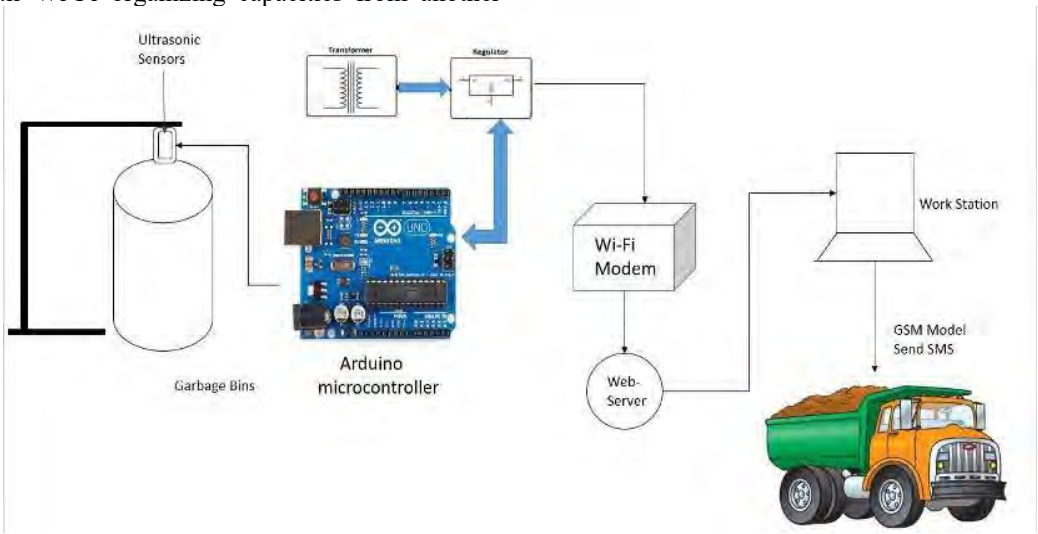


Fig.1: System Architecture

IV. HARDWARE USED:

Microcontroller: It get data from sensor and prepare on it. It compares the gotten information with the limit level set and appropriately yield is created. The microcontrollers are based on a 16/32-bit CPU with real-time imitating and executed follow back, that combine the microcontroller with 32 kB, 64 kB, 128 kB, 256 kB and 512 kB of executing high-speed streak memory.A128-bit wide memory interface and one of a kind quickening agent engineering empower 32-bit code execution at most extreme clock rate.[3]

Power Supply: We utilize 12v control supply in our venture. It is basically utilized to supply DC voltage to the components on board. 3.3V for lpc2138 and 4.2v for Bluetooth module is apply from control supply. 5V is required for transfer connected from control supply.[2]

Ultrasonic Sensor: The Ultrasonic Sensor sends out a high-frequency sound beat and after that times how long it takes for the resound of the sound to reflect back. The sensor has 2 openings on its front. One opening transmits ultrasonic waves, (like a modest speaker), the other gets them, (like a modest amplifier). The speed of sound is around 341 meters (1100 feet) per moment in discuss. The ultrasonic sensor employments this data together with the time distinction between sending and accepting the sound beat to decide the remove to a protest.

HC 05 Bluetooth Module: The HC 05 Bluetooth module is the foremost prevalent module within the Indian market and this module is generally utilized within the inserted ventures. The HC 05 Bluetooth modules are simple to utilize & straightforward, its cost is moo and these sorts of modules are interfaces with the Arduino, Raspberry Pi, and Microcontroller through the serial UART interface. These modules are outlined for the transparent remote association setup and it is exceptionally simple to utilize within the Bluetooth serial harbour protocol. The serial harbour, Bluetooth module is exceedingly qualified Bluetooth with the form of the V2.0+Enhanced Information Rate of 3Mbps. The tweak has totally 2.4 GHz radio handset and baseband. These seriaharbour, Bluetooth modules utilize the CSR Blue core 04-external single chip Bluetooth framework by utilizing the COMS innovation additionally with the Adaptive Frequency Bouncing Highlight. The impression of this Bluetooth module is as little as 12.7mmx27mm. Consequently it'll offer assistance within the by and large plan and advancement cycle.[5]

Servo Motor: A servo engine is an electrical gadget which can thrust or turn a question with incredible exactness. In the event that you need to turn and question at a few particular points or remove, at that point you utilize servo engine. It is fair made up of basic engine which run through servoinstrument. On the off chance that engine is utilized is DC fuelled at that point it is called DC servo engine, and in case it is AC fuelled engine at that point it is called AC servo motor. Able to get an awfully tall torque servo engine in a little and light weight bundles. Doe to these highlights they are being utilized in numerous applications like toy car, RC helicopters and planes, Mechanical autonomy, Machine etc.

Buzzer: We utilize a buzzer that's a critical component of our surveillance framework. Within the occasion a trespasser bypasses the manual security, our observation framework will not miss the trespasser. When identified close the restricted area, an alert is raised with the assistance

of the buzzer, which can alarm the security individuals and drive the trespasser to run away. We make utilize of a piezo buzzer, which is promptly accessible and cost-effective. A piezo buzzer is built of a piezo-electric fabric. The working of the piezo buzzer is based on the converse of the guideline of piezo power. The rule of piezo power states that at whatever point mechanical weight is connected on piezo electric fabric, it produces power. And piezo buzzers take after the inverse rule, which states that when a voltage is connected on a piezo electric component, it misshapes back and forward (a alter in measurement takes put) and produces an capable of being heard sound.

V. METHODOLOGY:

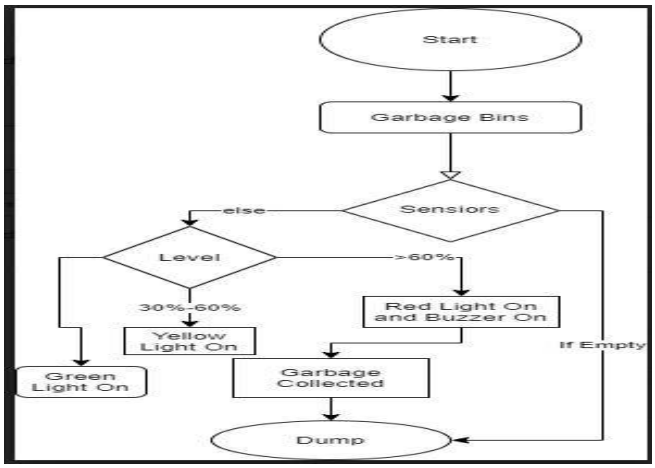


Fig.2: Flow Chart (Garbage Moniitoring)

ADVANTAGES:

Monitors the trash canisters and advises approximately the level of rubbish collected within the

the extraordinary conditions either full or empty. Some impediments of this observing framework are as follows:

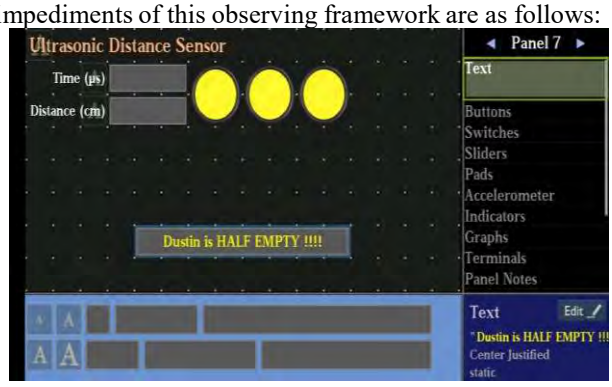


Fig.4: Bluetooth App Interface (Half Empty State)

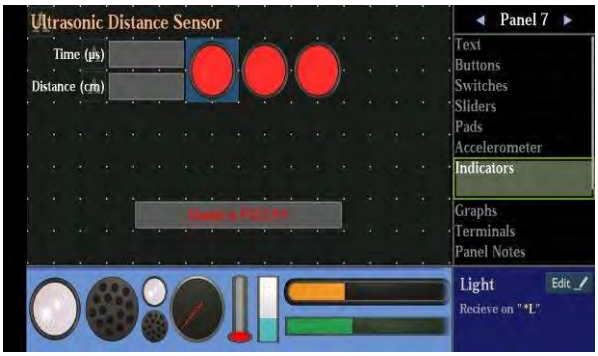


Fig.5: Bluetooth App Interface (Overflowing State)

VII. CONCLUSION

In this extend, a coordinates framework of Bluetooth, modem, IoT, Ultrasonic Sensor is presented for proficient and financial trash collection. The created framework gives made strides database for waste collection time and squander sum at each area. We dissected the arrangements as of now accessible for the execution of IoT. By actualizing this venture we'll maintain a strategic distance from over streaming of trash from the holder in private range which is previously either stacked physically or with the assistance of loaders in conventional trucks. It can consequently screen the trash level & send the information to collection truck. The innovations which are utilized within the proposed framework are great sufficient to guarantee the common sense and culminate for strong rubbish collection handle observing and administration for green environment.

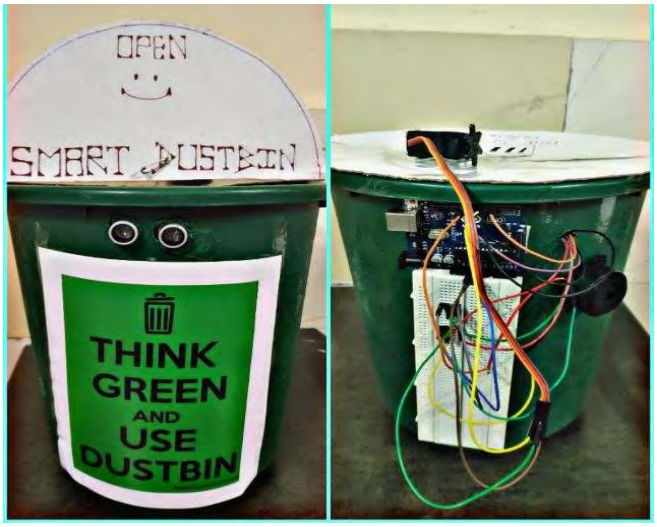


Fig.6: Working Model (Smart Dustbin)

References

1. Dipali Patil, Pradnya Gavhane, Priyesh Gharat and Prof.Urvashi Bhat, "IOT BASED Savvy Participation Framework Utilizing GSM", 7th Universal Conference on Later Patterns in Designing, Science & Administration,2017.
2. Dhvani Shah and Vinayak Bharadi, "IoT Based Biometrics Usage on Raspberry Pi", 7th Universal Conference on Communication, Computing and Virtualization, pp. 328 – 336, 2016.

3. Nikhil P. Shegokar, Kaustubh S. Jaipuria and Amitkumar Manekar, "REVIEW Mechanized Understudies Participation Administration Framework Utilizing RASPBERRY-PI AND NFC", Worldwide Diary of Inquire about in Computer & Data, vol. 1, 2015.
4. Mahalinga V.Mandi, Ashwini K.S., Chaitra H.S, Kavitha R and Kavitha U, "Biometric Based Participation Administration Framework Utilizing Wi-Fi", Universal Diary of Developing Innovation & Inquire about, vol. 1,2014.
5. Karthik Vignesh E, Shanmuganathan S, A.Sumithra, S.Kishore and P.Karthikeyan, "A Secure Biometric Participation Administration System", Universal Jour

Portable Electrocardiograph (EGC)

Anjali. P. Mishra, Neha. A. Jha, Tanya. R. Gupta
anjali301199@gmail.com, jhanehatcet@gmail.com, tanyaravikantgupta@gmail.com

Abstract— **IoT, i.e. Internet of Things, is the currently the most trending field in the Branch of Information Communication Technology. New medical technologies have arisen as a result of ageing population and an increasingly greedy demand for better healthcare. Some are focused on prevention, others in care. The system built here focuses on collecting electrical activity from the heart with a simple electrical circuit embedded in an Arduino microcontroller, with the aim of early detection of heart disease.An efficient and inexpensive approach was suggested, and the system performed hundreds of trials. These showed a high detection rate of R, S, and T-peaks (~100 percent, 78, 3 percent, and 72, 5 percent, respectively) with a specific measure of heart rate.**

Despite the good results achieved, this tool should not, for no w, be used for purposes other than monitoring. After studying the literature survey it was planned to overcome the research gaps and the system was tested using the test cases and then the outputs were generated. The output of the ECG were been shown according to different test cases and thus the Electrocardiograph become more responsive then those currently in use.

Keywords – Electrocardiograph(ECG); Internet of things(IoT); ADC Conversion; Arduino; Microcontroller

I. INTRODUCTION

Big efforts to provide good healthcare at affordable prices have become a central issue in the last decades. Along with wide technological developments, it has been possible to create more efficient devices and with new available features. Wearables are a category of devices that have grown exponentially in recent years, as battery life issues get overpassed. This electrocardiograph connects Arduino's features to an analog circuit and acquires, amplifies and filters a very low chest signal. A series of algorithms were then proposed to detect arrhythmias, cardiac muscle necrosis, hypertrophic ventricles, variations in cardiac rhythms depending on the position (down, sitting or standing) of the person. Then, a series of tests were conducted to determine the accuracy and reliability of the measurements obtained.

Despite having only been able to test our system in healthy people (thus, no pathological recordings were obtained), specific and representative results among all candidates were obtained. Given the results obtained against the simplicity of the circuit, it is predicted that significant changes will be made in a reduced time scale. Heart failure is generally a result of heart mechanical disturbances, but not electrical anomalies, and is not associated with specific patterns of the ECG. Nonetheless, some ECG results are encountered in cases of heart failure due to enlargement of the cardiac chamber and other related changes. Left atrial enlargement is known to cause characteristic changes in the P wave which sometimes show parallel path to the clinical condition. This objectively checks the ECG requirements for atrial enlargement. R wave height also showed first increase and then decrease in cases of cardiac enlargement, which was improved in a relatively short time period. Tissue edema was thought to be the main reason for the initial increase and the distance to the later decrease between the recording electrode and the heart, though other factors may contribute and change the findings.

Small disruptions in the conduction due to myocardial stretching and other causes may also lead to changes in the QRS complex. Examination of RR cycles provides a further possibility to describe heart failure. In cases of heart failure, a decrease in variability of the RR intervals and in high frequency components was recorded on frequency spectrum.

ECG is boon which helps to save many lives. The project focus on building an IoT based Portable Electrocardiograph using arduino as platform. Before the project starts there is the first chapter which speaks about the background of the project, why it is required and also it adds about the need and motivation for the project which is the recent incidents happened related also creating a great loss of life and and adds the scope of the project.Thus, the project was a great success with all the use cases and fulfilling all the phases of a project development and deployment.

II. LITERATURE SURVEY

A literature review collects scholarly publications, books, dissertations, conference proceedings, and other tools related to a particular issue, study area, or theory, and provides context for a dissertation through the analysis of past research. Research tells a story, and existing literature helps us figure out where we are in the story at the moment.

Continuing the story with new research and new perspectives is up to those writing a dissertation, but they must first be familiar with the story before they can move forward. The first piece of paper on A WEARABLE ECG MONITORING SYSTEM, based on IoT-CLOUD by the authors Zhe Yang ,Qihao Zhou , Lei Lei , Kan Zheng in December 2016. They invented and introduced a wearable monitoring system for electrocardiographs in this paper. The ECG data collected from the human body will be transmitted directly to the IoT cloud using Wi-Fi, without a mobile terminal being required. The outcome of this paper was to eliminate the need of mobile applications, the web-based GUI provides a versatile means independent of any mobile OS platform for users to access to the ECG data.

The second paper was written on A SURVEY ON HEALTH CARE APPLICATION Program IOT by the authors Sasippriya Saminathan, K.Geetha in 2017. In this

paper Body wireless sensor Network (BWSN) is used to transmit the patients' health parameters collected to the physicians and caregivers via Raspberry Pi microcontroller wirelessly . Being a long range wireless technology, the patient's health emergency situation is easily identified and prompt action leads to the patient's life as saved. This paper focuses on an all-encompassing health care monitoring system in real time using IoT and cloud computing service which are more beneficial for elders and chronic diseases' patients.

The third paper was written on An IoT-based low cost ECG Remote Patient Monitoring System by the authors Devendra R Sanghavi, S V. Athawale in 2018. In some nations, due to different obstacles people still do not have access to quality health facilities. The physical distance between the patient and the clinic is one of the main reasons for this. This paper explores the use of IoT in the area of health care and suggests a framework for tracking the ECG of distant patients. This system consists of Raspberry Pi, Arduino Uno, ECG Monitor and IoT Cloud for storing and plotting ECG data in real time. From this system it is seen that, the ECG monitoring system based on IoT is low cost and efficient for remote patient.

The next paper was written on Electrocardiography Circuit Design in 2013 by Nathan M Kesto . This application note will attempt to provide the reader with information on ECG signals as well the methods and design techniques used to construct an ECG demonstration board. It will focus on amplifiers for small ECG signals as well as on some of the different ways to reduce the various noises in the system; this includes explaining the Right Leg Drive circuit.

A Paper on a real-time ECG monitoring system based on IoT using Cypress WICED in 2017 by Uttam U. Deshpande, Milan A. Kulkarni. Monitoring of ECG based on Cypress Wireless Internet Connectivity for Embedded Devices (WICED). The ECG data was obtained via a wearable monitoring node and transmitted directly via Wi-Fi to the IoT cloud. The ECG signal gets distorted from sources such as ambient electrical interference, measurement (or electrode contact) noise, electromyogram noise (muscle contraction), movement artifacts, instrument noise (such as ADC conversion artifacts).

III. PHASES OF PROJECT

Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: Getting clear idea of the project title and doing research on it .So that we will get our definition and after that we will first create the Literature Survey of the project and do the whole documentation. Planning: After analysis we will first study about it and do some research on it for our better understanding of the project and also getting a rough picture about what would be our problem definition for the particular project.

Designing: Then we will draw the circuit diagram of the device and according to that, will list down all the requirements needed for the construction for the prototype of our project.

Implementation: After acquiring the requirements we first developed the connections of the hardware components(Arduino,Tiga-Mega Gold Electro deetc.)and then added the coding part later with help of Arduino data cable to the Arduino.

Phase 2: Testing andDeployment.

Testing:After the prototype is ready we will first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding to it and will check again.

IV. PROJECT PLANNING (RESOURCES, TOOLS USED, ETC)

Hardware Tools :

Quantity	Hardware Components
1	ARDUINO MEGA 2560, rev3, CODE:A000067
3	Tiga-Med ECG Solid Gel Electrodes 45 mm by EKG Einmalelektroden Schaumstoff
1	ECG Module AD8232
10	Connecting Wires

Software Tools:

1. Arduino IDE

Components:

Arduino uno

The Arduino Uno is an open-source microcontroller board based on the Microchip ATmega328P microcontroller developed by Arduino.cc. The board is fitted with digital and analog input / output (I / O) pin sets that can be interfaced with different boards of expansion (shields) and other circuits.The board has 14 Digital pins, 6 Analog pins, and can be programmed with the Arduino IDE (Integrated Development Environment) via a Type B USB cable. It can be powered by a USB cable or an external 9-volt battery, although it supports voltages from 7 to 20 volts.

Fig 3.1:Arduino Uno

Tiga Mega Gold Electrode

Tiga Mega Gold Electrode Spinal cord injuries are among the most traumatic situations, having relevant repercussions on an individual's occupation performance. Although the most significant effect is the loss of function, neuropathic pain can determine a person's inability to return to day-to-day activities. Hence, the development of new technologies with significant impact on the process of rehabilitation of spinal cord injury is important.

Fig 3.2 Tiga Mega Gold Electrode

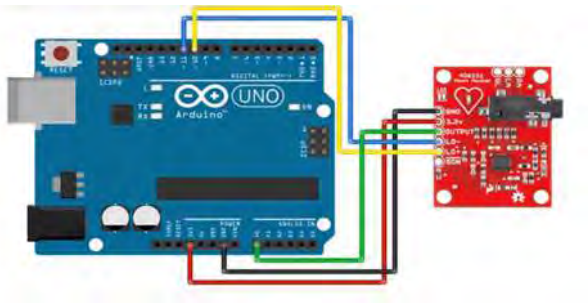
ECG Module

The AD8232 is a sleek little chip used to monitor the heart's electrical activity.

V. Design and Implementation

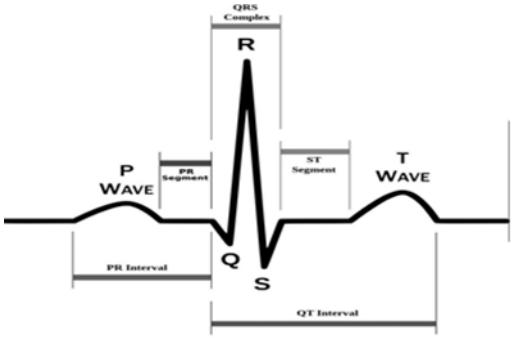
- Connected 3 electrodes to the body(1 on left arm,1 on right arm,1 on right leg).
- Connect the connector wires onto the electrodes.(Yellow to the left arm, Red to the right arm and Green to the right leg).
- Connect the AD8232 to the Arduino using jumper wires.
- Connect the Arduino to laptop and with the help of Arduino IDE upload the code to the Arduino.
- Once uploaded, check the serial plotter to see your heartbeat.

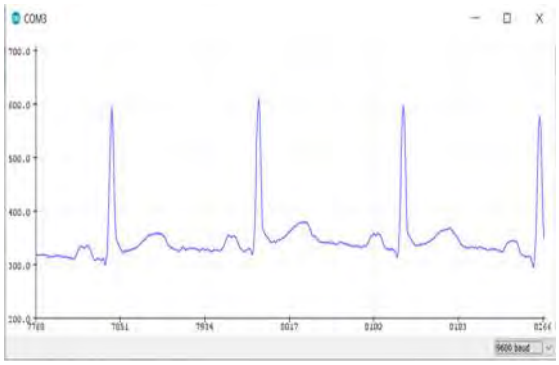
Circuit Diagram:



VI. Results

The normal one cycle ECG pulse output figure from a heart beat. From the figure one cycle consists of wave P, wave QRS up to wave T. P wave offers gain knowledge about the time of the impulse propagation to both atria. Then proceed with a flat pattern called PR segment arising from the transmission of the electrical impulse from atria to ventricles.And proceed with the complex QRS wave which can be seen in the figure above. The complex of Q, R and S consists of three small waves, i.e. a low Q wave, a high R wave and a weak S wave. The QRS network provides information on fibrillation and arrhythmias, and monitoring heart attacks can be helpful.





And then the ST phase, it follows the S wave and may point to the incidence of ischemia even with the T wave. This reflects the time during which ventricles contract, which is the final stage of the cardiac cycle. The T wave helps one to have cardiac hypertrophy information, heart attack and ischemia information. Certain parameters, such as the QT interval, also allow the characterisation of different additional pathologies. The ECG signal finally ended with a slight peak, U wave. Keep still while calculating. Normal movement of the muscle still produces the electric pulse and you really want to get rid of it. Place the pad near soft tissue, as electric pulse due to its poor conductivity can hardly pass through the skeleton.

VII. IMPROVEMENTS IN RESEARCH GAPS(FINDINGS)

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. They are as follows:

- When an irregular pattern is shown by the ECG system, multiple unrelated reasons can occur including a standard version. To sort things out, a doctor needs to do more thorough examinations, including other tests (e.g. echocardiogram).
- The biggest concern with ECG is probably false negatives. The EKG may be completely normal for some heart patients but their symptoms should be expressed in the ECG. The reasoning behind that is not well understood. A good reading of the ECG does not exclude the underlying heart disease and other symptoms, such as chest pain, need to be considered and further examination may be needed.
- Not all heart problems present on an ECG.Vulnerable plaque (a type of atheroma) is a prime example. Vulnerable plaque is a rapidly growing deposition or degenerative accumulation of plaques containing lipids on the innermost layer of an artery wall .Since artery walls usually expand in response to expanding plaques, they do not affect blood flow and can't be detected even in an ECG check for heart stress.But fragile plaque is a significant cause of heart attacks

VIII. FUTURE SCOPE

The future scope for Portable Electrocardiograph are:

1.Using IOT Technology (Through SMS Warning and Mail), data can be uploaded or sent directly to doctors so that everybody will be registered.

2. Measurement of the electrodes can be more precise by using more.

3. It can be used by normal humans as it is portable.

IX. CONCLUSION

A Portable Electrocardiograph was designed and developed. We were able to monitor the heartbeat of human . We obtained a good-looking signal with a very simple circuit embedded in a highly portable system. Staying still during measurement may give you accurate results. Placing the **Tiga Mega Gold Electrode pads** near soft tissue can provide accurate results since electric pulse hardly travel through skeleton due to its poor conductivity. Future studies related to ECG monitoring are still needed in the future. It is believed that long-term and user-friendly ECG monitoring can greatly help mollify existing healthcare problems to a convinced matter.

References

[1] IEEE, Kan Zheng, Senior Member, IEEE, Wei Xiang†, Senior Member, IEEE- *An IoT-cloud Based Wearable ECG Monitoring System for Smart Healthcare*|Journal of Medical Systems · December 2016|

[2] Devendra R Sanghavi, S V. Athawale-*An IoT based Low-Cost ECG Monitoring System for Remote Patient* | Volume 3, Issue 4 | ISSN: 2456-3315|

[3] 1Sasippriya Saminathan, K.Geetha- *A SURVEY ON HEALTH CARE MONITORING SYSTEM USING IOT* |Volume 117 No. 17 2017, 249-254| ISSN: 1311-8080|

[4] Nathan M *Kesto-Electrocardiography Circuit Design* |4/5/2013|

[5] Uttam U. Deshpande, Milan A. Kulkarni-IoT based Real Time ECG Monitoring System using Cypress WICED |Vol. 6, Issue 2, February 2017|ISSN : 2320 – 37

[6] P. Singh and A. Jasuja, "IoT based low-cost distant patient ECG monitoring system." Published In International

[7] Conference on Computing, Communication and Automation (ICCCA), Greater Noida, 2017.

[8] [2] M. R. F. Nurdin, S. Hadiyoso and A. Rizal,"A low-cost Internet of Things (IoT) system for Multi-patient ECG's monitoring." International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), Bandung

[9] [3] Lin, Alice M, Wong and Kevin C, Tseng et al. "Community-Based ECG Monitoring System for Patients with Cardiovascular Diseases." Journal of medical Systems, 2016.

[10] [4] M. Neyja, S. Mumtaz, K. M. S. Huq, et al., "An IoT-Based E-Health Monitoring System Using ECG Signal."

[11] GLOBECOM 2017 IEEE Global Communications Conference, Singapore, 2017.

[12] E. Span, S. Di Pascoli and G. Iannaccone, "Low-Power Wearable ECG Monitoring System for Multiple-Patient Remote Monitoring," in IEEE Sensors Journal, July 1, 2016.

Automatic Medicine Reminder

Nevil Kaspale, Shubham Maheshwari, Rajeshkumar Jha
Thakur College of Engineering and Technology
Mumbai, India
tcet@thakureducation.org

Abstract--**The idea of a digital world where different types of sensors and common processing among them to slice information is widely used in many industries nowadays. There are a lot of products which are developed using these ideas. Healthcare industry is the one where there are a lot of improvements already taking place. Medicines play a major role in every individual for the prevention and cure for most diseases. Many harmful diseases can be solved through proper medication time. There are a lot of people who take their medicine regularly. Our Project's main aim is to make a Smart Medicine Reminder system using NodeMCU for users who take medicines regularly for longer periods as one often tends to forget one or the other dosage. It is very important for such users to take the medication as per scheduled by the doctor for their care. Also, for the patients who are suffering from the permanent diseases like diabetes, blood pressure, breathing problem, heart problems etc. forgets to take the pill on proper time. Similar kinds of problems are also observed in hospitals where nurses monitor the patients and its quite difficult for them to keep the track of the medicines of each and every individual. In this Project, a medicine reminder system is designed which reminds the user to take the medicine once or twice or thrice a day. The time slots can be selected using push buttons. Also, it shows current Date and time. The medication reminder can also be integrated with Patient monitoring System.**

Keywords – NodeMCU; Automatic Reminder; RTC Module; Blynk;Arduino;

I. INTRODUCTION

In our modern, day-to-day busy lives, the number of people consuming medicine has increased drastically. The reason behind this is diseases are increasing in large amount this was not the case few years back. Poor medicine adherence despite everything is a significant urgent looking by a large portion of industrialized nations prompting declining illness seriousness and expanded expense related with higher hospital admission rates[4]. To overcome our illnesses and live a better life we need to take medicines regularly. One should be in counsel of the Doctor who instructs us to take wanted pills in wanted manner with the goal that

patients face issues like overlooking pills to take at ideal time and furthermore when Doctor changes the medicine of medication patients need to recall the new schedule of medication[5]. Forgetting pills at right time, taking wrong medicines and also accidentally taking expired medicine is a major problem faced by many people and this leads to them suffering from an unhealthy life. Our project is to make a NodeMCU based smart medicine reminder which uses real time clock and gives an audio-visual alert for the given medicine type.

II. BACKGROUND

There are many people who constantly need help be it the elder ones or our family members who needs some special needs. With aging, people experience poor eyesight and poor memory. It may happen that some people may forget to take medicine at the correct time and forget the medicine which they have to take. In order to eliminate the factors of always needing observation or taking a risk of missed doses one must find an easy, portable, efficient and smart solution. . In order to make a very useful and smart solution it had to be integrated with smart technologies. While at the same time it had be fit for the elders and their limited knowledge and experience to implement the ease of use. Size and portability were also an important fact that we had to keep in mind. For it to be called smart, it's connected through a wireless network, which enables it to be connected to the internet for the future applications and integration. Through that equivalent system its associated with the cell phone, which with it you can set the planning interim for the portion and furthermore informs you by numerous ways when the portion opportunity arrives. Also, we added a notification system such that it alerts you on your mobile phones via notification and an Email. Since nowadays phones are a very handy product, so it seems best to act as a medicine reminder.

Importance of the project

With increase in the number of people suffering from diseases it is very important for them to take medicines on a regular basis. Smart medicine reminder helps one to take their medicines on time. Medicine reminder is very wide and can be used by patients at home, doctors at hospitals and at many other places. Importance of the medicine reminder using NodeMCU on day-to-day life is as below:

- Audio alert reminds the patient about the medicine.
- SMS alert is also important for the people whose daily life schedule is very busy and to remind them about the medicine, SMS alert is very effective.
- The proposed system is very useful in hospitals where a nurse monitors more than 4- 5 patients about the reminder of medicines.

As a human being almost, everyone has the habit of forgetting things and taking thing very lightly but it can cause a lot of damage sometimes. Considering the scenario where a nurse monitors many patients it may happen that they might forget some medicines for some patient and can cause a severe damage if that patient is in a critical situation and also for the people who are suffering from the permanent diseases like diabetes, cancer etc. It's very important for them to take medicines on a daily basis to treat the symptoms of the disease. It's even more important for the older people suffering from

these diseases to take medicines on a regular basis. With the world moving towards digitization and using smart products it became a very challenging aspect to solve the above-mentioned problems and it gave the motivation to help something for the society and we decided to develop a simple cost effective and smart medicine reminder using NodeMCU.

Objectives and scope of the project

This aim of this project is to build a smart medicine reminder for the people who tend to forget to take the medicines on time. When the pill time has been set, the system will remind patients to take pills utilizing audio and visual alerts. This model can aid elders to take their medication[6]. The proposed system can be extended to patient monitor system and can be used for continuously monitoring patients. Moreover, it can be used by hospitals to reduce human efforts.

The proposed system is designed for, but not restricted to, helping old people in taking care of themselves in taking their medication at the correct time and in the correct amount[6]. It has been observed that people in general neglect their health and also give preference to the other things than taking their medicines. This is also the reason they forget to take their prescriptions on time. Many health maintenance organizations, health practitioners and medical researchers have realized that increased use of patient reminders can significantly increase the treatment of chronic illness and delivery of medical services to the patients who need it.

III. LITERATURE SURVEY

In this method, an Android based application is created for the patients. This application will remind their customer to take the required medication at an appropriate time by setting the updates in a compact and systematic manner. These updates can be set in the application in total accordance with the patient's prescription. Urvashi Sharma, Chetna Chauhan, Himani Sharma, Anjali Sharma had proposed a Smart Medicine Reminder using Arduino. ARDUINO module comprises a micro-controller which controls the function of the system and performs all the tasks[1]. No external peripheral devices and memory devices are required. Many systems serving the same purpose are difficult to use, non-mobile, expensive and complex processes. The proposed system overcomes these problems. Priyadarshini, Ramya, Kalaiyarasi, have inquired about a novel approach of micro-controller based Automatic Medication Reminder (AMR) structure for patients. In this approach gear worked using microcontroller is used for the patients. This application will give a refresh using ringer and LCD will demonstrate the name of pharmaceutical. This structure is insignificant as it relies upon microcontroller which is difficult to embed.

Aditya Vijay and Durgesh Kumar have proposed a system built on customer server correspondence where Client specified information will be stored at server end in a MySQL database and USB link is used to refresh information into the Arduino[2]. The device alerts via an alarm beep and a message on the LCS successfully. But this framework is hard to utilize, non-portable, costly and complex.

Tahseen Fatima, Pallavi B V, Dr. Baswaraj Gadgay's proposed device consists of a small box divided into seven compartments for storing pills for a week, each compartment is sub divided into 3 sections[3]. An Arduino mega 2560 is used, which processes the activities and accordingly displays the pill details and time of intake on the LCD attached to the box and a GSM module which sends message to the family physician or members in case the pill is taken or not taken and also sends message about patient health condition. The device helps in keeping track of regular medicine taking activities and reduces manual supervision and human effort.

Sanjay Bhati, Harshid Soni, Vijayrajsinh Zala, Parth Vyas, Mr. Yash Sharma have proposed a Smart Reminder Box in which present time will be saved in RTC module and notification time will be saved in EEPROM[5]. At the time of taking medicine system generate Notification sound and display the Bright light in certain pill boxes. It can sense if the patient had taken out pills from the box or not. This system is useful for especially those people who are taking pills regularly, prescription of medicine is very long and hard to remember for those users. The proposed system is also reusable by exchanging those other medicine box that has only alerting system and are non-usable or affordability compare to our product.

IV. PLANNING AND RESOURCES

Project planning (Resources, Tools used, etc.)

Hardware Tools:

1. NodeMCU
2. RTC Module
3. Breadboard Connecting wires

Software Tools:

1. Blynk
2. Arduino IDE

NodeMCU

NodeMCU is an open source Lua based firmware for the ESP8266 WIFI SOC from Espressif and uses an on-module flash-based SPIFFS file system. NodeMCU is implemented in C Programming Language and is layered on the Espressif NON-OS SDK. The firmware was initially developed as is a companion project to the popular ESP8266-based NodeMCU development modules, but the project is now community- supported, and the firmware can now be run on any ESP module.

RTC Module

Real Time Clock(RTC) is an Integrated Circuit(IC) which is used as timekeeping device. RTC is an important component of many time critical devices and applications

like Servers, GPS, and Data Loggers etc. The DS3231 is an RTC IC developed by Maxim Integrated. It is a comparatively cheap, and extremely accurate RTC IC with communication over a I2C Interface. An interesting feature of the same is that it comprises an integrated crystal oscillator and a temperature sensor, hence eliminating the need to connect an external crystal.

Blynk:

-Blynk App => allows to you create amazing interfaces for your projects using various widgets we provide.

-Blynk Server => responsible for all the communications between the smartphone and hardware. Blynk Cloud can be used or you can run your private Blynk server locally. It is open source, can easily handle thousands of devices and can even be launched via a Raspberry Pi.

-Blynk Libraries => for all the popular hardware platforms - enable communication with the server and process all the incoming and outgoing commands.

In this after understanding the topic the project feasibility was analysed by performing different types of feasibility studies and by also planning the project tools. Feasibility study will help in better understanding the various feasibilities associated with the project and helping to make the correct decisions and completing the project within the schedule, budget, etc. The tools were specifically identified in this chapter in accordance with which technology can be feasible and how efficiently can the project be implemented. This helps us in understanding the tools and technology that can be used for the project.

IV. PROPOSED SYSTEM

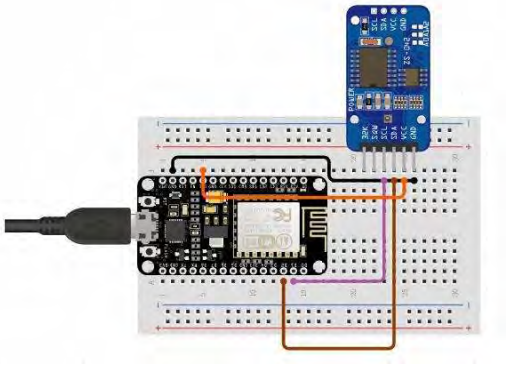


Fig 1 Block diagram

The NodeMCU can be connected to any power supply source. In our case a laptop powered the NodeMCU module. Then we specify a schedule in Arduino IDE and run the program. The RTC or Real Time Clock counts seconds in real time and sends to the NodeMCU. The Blynk App is wirelessly connected to the NodeMCU. Our program is such that if the specified time matches the real time then a notification will be sent to the user in addition with an E-Mail Reminder for the same via the Blynk App.

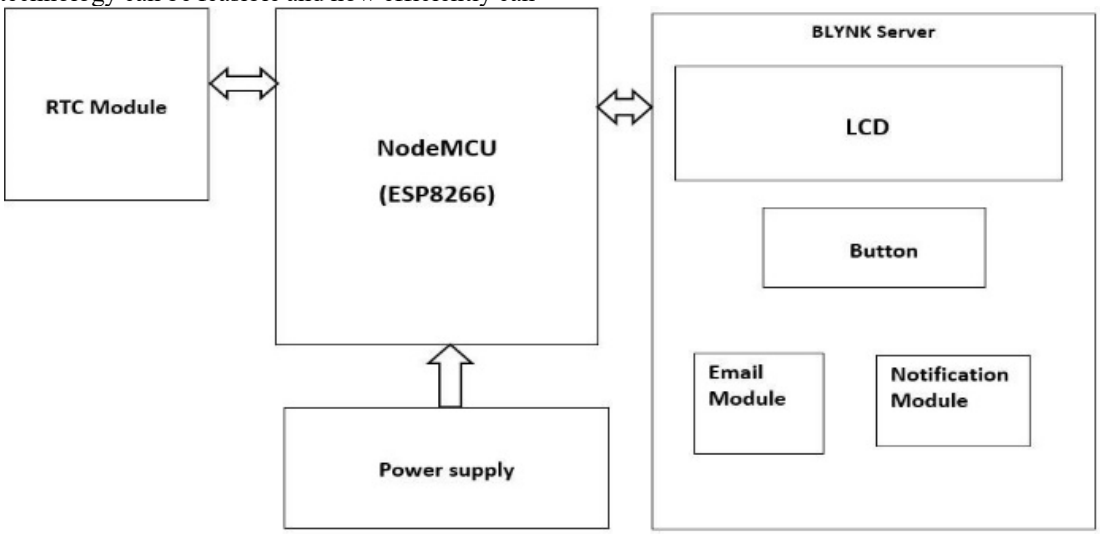


Fig 2 Circuit Diagram

V. RESULT AND DISCUSSION

1. When the time set on the device a notification as well as an email will be sent to the recipient.

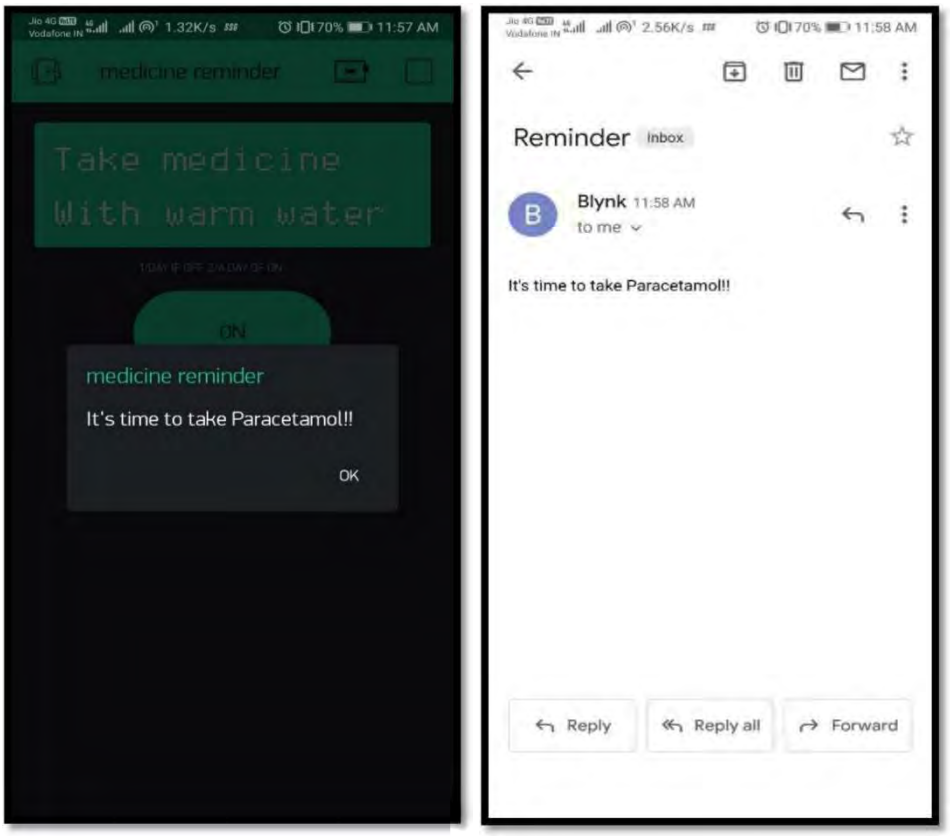


Fig. 3 Mobile and Email Notifications

2. Even if the recipient is offline, they will receive a notification

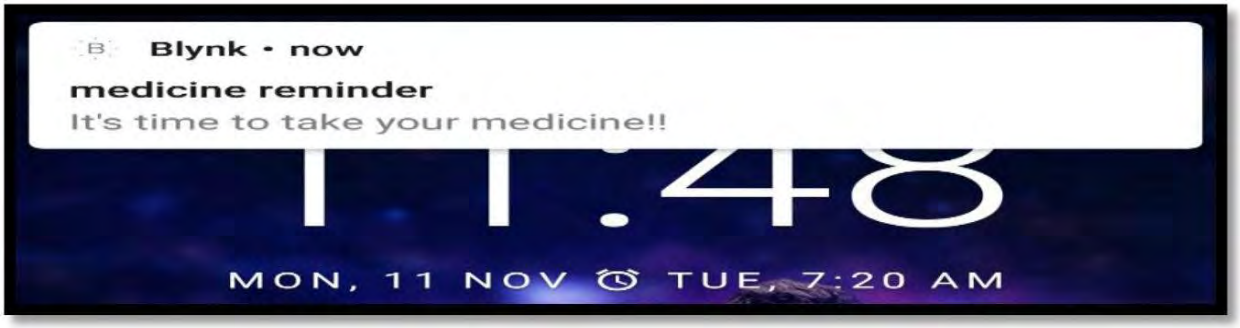


Fig. 4 Mobile Notification

Discussion :

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. As in this system the timing for the reminder was set in the code itself, in the next version the time can be set manually. Also, this system does not allow user to set number of days for which they have to take the medication, we can implement it to make their life easy. A push notification regarding the amount of smoke can also be sent to the owner and their neighbours and contacts. If the amount of smoke is high, a notification can be sent to the concerned nearest emergency contacts, such as the nearest fire station, along with the entire location of smoke site. The biggest concern is the size of this system, we can try to make it small in size so it can be portable and easy to use.

VI. CONCLUSION

There are numerous frameworks which are filling for a similar need. Be that as it may, these frameworks are hard to utilize, non-portable, costly and complex process. The proposed framework defeats these issues. The NodeMCU Based Medicine Reminder is easy to utilize, moderate, better exactness. This framework is useful for each age gathering and can likewise be utilized as a part of healing facility for a gathering of individuals. This framework will decrease the awful impact caused because of wrong admission of medication. In hospitals, a screen can be placed that is connected to the system, outside their ward to notify the nurse to give patients their medicine. Moreover, it can be used by the families that have elderly people that always forget to take their medicines.

Currently, there are many medication reminder system which is operable manually. Due to manual work, the

available system becomes more time consuming. So, in the given work, an attempt has been made to implement fully automatic medication reminder system. The system reminds user about their medicine intake schedules.

VII. ACKNOWLEDGEMENT

The author would like to thank subject lecturers current and previous), curriculum planners and other computer and mathematics school staff members who have previously planned, created and taught this subject at different times

References

[1] Urvashi Sharma, Chetna Chauhan, Himani Sharma, Anjali Sharma, "Arduino Based Medicine Reminder", AGU International Journal of Engineering and Technology (AGUIJET), Volume No. 3, e-ISSN: 2455-0442, p-ISSN: 2455-6734, July-December-2016.

[2] Aditya Vijay, Durgesh Kumar, "Automatic Reminder using Arduino", International Journal of Innovative Research in Technology (IJIRT), Volume 4 Issue 11, ISSN : 2349-6002, April 2018.

[3] Tahseen Fatima, Pallavi B V, Dr. Baswaraj Gadgay, "Arduino Mega Based Pill Reminder", International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue VII, July 2018.

[4] Pallavi Prakash Ghogare, Mrs. V. S. Jahagirdar, "Medicine Reminder and Monitoring using IoT", International Research Journal of Engineering and Technology (IRJET), Volume : 05, Issue : 04, e-ISSN: 2395-0056, p-ISSN: 2395-0072, April-2018.

[5] Sanjay Bhati, Harshid Soni, Vijayrajsinh Zala, Parth Vyas, Mr. Yash Sharma, "Smart Medicine Reminder Box", IJSTE-International Journal of Science Technology & Engineering, Volume 3, Issue 10, ISSN:2349-784X, April 2017.

[6] Diaa Salama Abdul Minaam, Mohamed Abd-Elfattah, "Smart drugs: Improving healthcare using Smart Pill Box for Medicine Reminder and Monitoring System", Future Computing and Informatics Journal 3, 443-456, November 2018

Automated Pet Feeder using Arduino

Shubham Mishra, Lokesh Mishra, Pragya Khandelwal

TCET- Mumbai

shubhi.mishra5555@gmail.com, lokeshmishra2@gmail.com, pragya.khandelwal2@gmail.com

Abstract— **over the years, more and more households are beginning to have pets. Pet feeders came into existence as more and more pet owners found it difficult to cater time to feed their pets. Pet feeders are automated machines that dispense food at preset schedule. They are mainly timed based and dispense a certain amount of food at specific time of the day. The pet feeder is a programmable system which is mainly controlled by a microcontroller. It consists of a LCD screen for input display, buzzer to alert pets for meals, stepper motor to control the speed.**

The work is about pet feeding machine automatically for a daily minimum period of time of eight hours when all people of nuclear family members are busy at work for the survival in the metro cities and other cities. This set up is controlled by the Mobile app automatically. The pets of home can get food after an interval of three hours and the same can be monitored using mobile app and the owner of the pet always ensure about the feeding of the pet especially dogs and cats when they are busy at work. The food for pet has been kept inside a box while leaving home. The food would be served to the pets automatically up to a certain quantity only when the pet comes near to the box when they feel hungry. It is an excellent idea to protect the pets from starving whenever no one is at home and all are busy with their scheduled routine works and job at office or in business.

Keywords – Pet feeder; LCD screen; Mobile Application; Arduino Mega

I. INTRODUCTION

Dog is the most common pet that people owned. Nowadays dog is also one of family member to the owner. So, dog's healthiness is important. Regular feeding is the most important thing, but it also one of the problems in dog's maintenance. Owners often forget to feed their pets because of their work. A dog feeder can be the answer to these problems. For this project we will design and build a prototype of an automated pet feeder which will be suitable for use by cats and small dogs. This product will be called the Smart Pet Feeder. This system will hold enough food for 5 feedings. This device can provide regular feeding without disrupting owner's work. Owners can monitor feeding process with their Android smartphone virtually. Smart Dog Feeder can give authentication with RFID, set feeding time and portion per serving through Android smartphone, send feeding report (eaten or partially eaten), and dog arrival when the feeding time has arrived. Every setting about feeding time, portion, stock and waiting time will be set on Android App. Smart Dog Feeder will save the schedule and set alarm which will interrupt when the feed time has arrived. Food will be served based on user's setting and be measured by load cell. Experiment is done by seeing punctuality, portion congruence, delivery of settings and notification within devices. The result of experiment is Smart Dog Feeder can receive messages from server and do feeding at the right time. The Smart Pet feeder will give pet owners a solution to this problem, thereby improving the lives of both pets and the owners.

"Technology stems from humanity." This sentence is a good description of modern life. The interaction between humans and physical devices and objects in the real world is gaining more attention and requires a natural and intuitive methodology to employ. Cyber physical systems refer to a new generation of systems that integrate computational and physical capabilities and are capable of interacting with humans through many new modalities. Thus, this study examines the ability of computation, communication, and control technologies to improve human interaction with pets.

Different sensors are used for the automatic pet feeder so that it works efficiently. A distance sensor or a proximity sensor will be connected to the Arduino and as the pet is sensed in the vicinity of the pet feeder, the food from the bottle will be poured in the bowl. Whenever the distance sensor senses motion at a particular distance from the pet feeder i.e. as the pet comes closer to the bowl, the food is served. A servo motor will be used for the locking system. It will be similar to the locking system controlled by angle. All these components will together determine an efficiently running automatic pet feeder. The cost of a care taking of a pet has been reduced down due to the launching of such machine in the market. Such idea is being globally getting popular due to Automatic system and mobile app involvement.

Automatic pet feeding system features a machine which can feed pets (e.g. dogs) automatically after frequent intervals in absence of his master. By using machine master don't have to stay with his pet every time to feed it and he gets liberty to do his other works outside without caring about his pet. Dish for feeding pet could be filled in number of ways one is we can set the time and date using Arduino UNO which is displayed on LCD screen fitted on body of pet feeder.

The effectiveness of any application also depends on how the products are detailed and how the components are used. The most important part is of the ultrasonic sensors to detect the presence of the pet because if it fails to work then it is not possible that the model can work. The cost is also low as per our model requirements.

II. SUBJECT HISTORICAL DATA

In this section, we present the previous research done by the people by reading their papers published. Also, the problem definition, features of the project and methodology is explained and learned in detail.

A. Literature Survey

Mritun Jay, Subhash Chandra Tiwari, Sahil Manoj Hawal in their journal "International Journal of Innovative Research in Science ,Engineering and Technology(2018)" states that it was based on Arduino. They have used various sensors for various purposes. They have provided real time notification, feedback on web-server in which customer can see what is happening in their home. Rachel Heil, Kristine McCarthy in their book "WENTWORTH INSTITUTE OF TECHNOLOGY(2008)" states that the proposed system makes sure that each pet has access to a healthy amount of food throughout the day, regardless of the owner's schedule and each pet eats only its own food .

B. Problem Definition

Dog's care should be enjoyable not burdensome and so the goal of this project is to assist owner with pet care by providing an automatic pet feeder IOT based system. The purpose of the project helps the owner of the pet feeding their pet on time even when they are not at home. Other than that, it also can help the owner know the diet of their pet. Knowing the diet of the pet is very important for the owner to make sure that the pet is in good health. This system assists pet owner to feed the pet. The system act in two ways, one is feeding the pet and sends the feeding information to owner. Sick pets require special diets and large amounts of time and money and is difficult to make sure the correct pet is receiving medication or special food. No product on the market to addresses this issue which made us develop this automated feeding system which makes it easier for the owner.

C. Features of the Project

The proposed system is an automated Pet feeding system, consists of sensors. Server controls and monitors the various sensors, and can be easily configured to handle more hardware interface module (sensors).. Wi-Fi technology is selected to be the network infrastructure that connects server and the sensors. Wi-Fi is chosen to improve system security (by using secure Wi-Fi connection), and to increase system mobility and scalability.

The features of our project can be highlighted in following points:

- •Simplified automation.
- •Minimal controls
- •Cost efficient
- •Can be upgraded if needed

D. Methodology Used

In today's world everyone has shortage of time and in this era of technology there should be a mobile app that could be used to control the home appliances with the help of speech recognition.

This project will use the system which will be powered with Wi-Fi, making interconnectivity between the devices and creating an internet of things. In this project the Arduino Mega 2560 and the Arduino Wi-Fi shield have been used to implement the smart pet feeder.

- •In this proposed project a mobile app is used which includes all the features of controlling the feeding system with the help of interconnectivity of devices.
- •The mobile app contains all the commands like turn off, turn on, set timer etc.
- •The user is authenticated using a unique authentication token generated by the server, which is then hardcoded into the microcontroller.
- •After successful authentication the user will be able to control all the appliances with the mobile app.
- •No longer need to worry about their pets during business trips or vacations.
- •No longer need to purchase multiple feeder for multiple pets.

III. ANALYSIS AND PLANNING

Analysis is an important part of any project to know calculate the wrongs and rights of the project i.e. the economic study tells us the capital used, the planning tells us about the time taken, resources used gives us all idea about how well the project is being going on.

A. Fesiability Study

Analysis is an important part of any project to know calculate the wrongs and rights of the project i.e. the economic study tells us the capital used, the planning tells us about the time taken, resources used gives us all idea about how well the project is being going on.

1. Economic Feasibility:

This project is based on Android App and few electronic components like Arduino mega 2560 ,relay switches etc. which are affordable, making it economically feasible to implement.

2. Technical Feasibility:

This project is based on wireless technology and embedded system which is reasonably in phase with currently used technology. Therefore, it is very much favored by the technology.

3. Operational Feasibility:

This software will have very easy to use, user friendly interface so it will be pretty much operable by anyone having little experience of using an Android Device. It could be helpful for physically disabled person too, controlling home appliances with the click of a button. So, it is operationally feasible.

B. Assumption Risk:

a) Assumption:

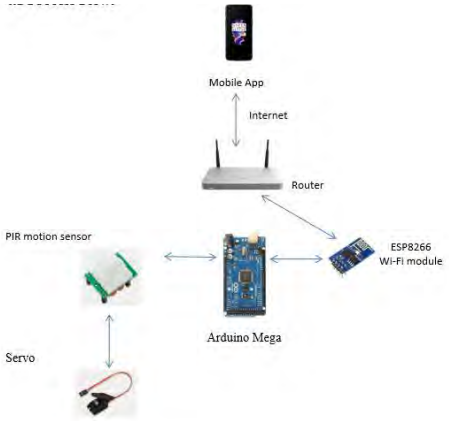
- Management will ensure that project team members are available as needed to complete project task and objectives.
- The team members will participate in the timely execution of the Project Plan (i.e., timely approval cycles and meetings when required.)

b) Risks:

- Accurate identification of the system requirements, interpretation and design depend on the patience and commitment of the team members in focus.

IV. DESIGN REQUIREMENTS

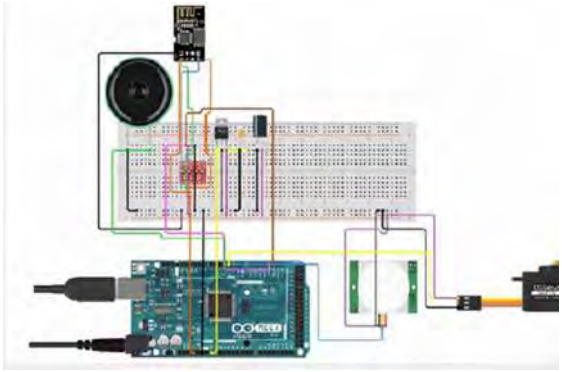
C. Process Flow



This is the process flow of the project.

D. Block Diagram

The following is the Breadboard implementation of the project which can also be converted to a PCB in order to make the assembly smaller.



V. RESULT

Results show that everything worked as anticipated, although some high pings in the network may cause minor occasional lag in the connection and data exchange, however, there was no major lag detected affecting the model control and testing. The lag maybe a casual effect of using 3G connection, however, most Pet Feeder systems employ DSL or Broadband connection, resulting in acceptable performance. Eventually, expected results were met. Automatic Pet Feeding System was planned to ensure the time to time feeding of pet in absence of its master so that master can do his other tasks without worrying about feeding. Automatic Pet Feeding System has attractive design and aesthetic model. Arduino and IoT add Automation in the system. The report showcased the basic design of the system to be made. And also the Arduino circuit to control the functions of the system. Automatic Pet Feeder works efficiently and fulfils the objective of feeding pet in absence of its master. It works for on house hold 230V AC supply. The servomotor rotates the Auger and food gets delivered into plate as programmed in the Arduino. On the backend, there is Arduino mega 2560 which has been programmed to control the status of GPIO pins and thereby controlling the appliances.

VI. CONCLUSION

This project brought together several components and ideas to achieve a common goal that is design an automated pet feeder using Arduino mega. The key components of the project include a distance sensor to detect or identify the presence of the pet i.e. whether the pet is near the bowl or not and a servo motor which will be programmed to serve the food as soon as the pet comes closer. It relieves the owner from having to feed his pet multiple times a day. The proposed project senses the presence of the pet using the distance sensor and serves accordingly. The owner does not have to worry about making plans or feeding his pet because of this automated pet feeder. This automatic pet feeder serves as a helping hand as it works efficiently in the absence of the owner.

VII. FUTURE SCOPE

IoT is a platform which can embed both software and hardware. It is obvious from that IoT is an efficient way to access data. As suggested, SOAP based mechanism with web services is an ideal choice for managing diversified devices and appliances in home environment. Various sensors are used to monitor various activities of the pet say, an IR sensor is used to monitor if food is available in plate or not. The design can be made more factories feasible and aesthetic. The auger can be made by 3D printing which can be created as the auger is the moving part. The APF could be made more advanced by installing cameras and audio box as it would make possible to interact with pet through smart phones as well as keep observation on it. APF is indeed a helper to the owner for proper care taking of pet and more advancement can be expected in future.

VIII. ACKNOWLEDGEMENT

We sincerely thank our guide Mr. Anil Vasoya for his guidance and support for carrying out our project work. We also thank all the project coordinators for arranging the necessary facilities to carry out the project work.

We thank our Principal, Dr. B.K Mishra, HOD, Mr. Rajesh Bansode and the College Management for their support. We would like to thank all our classmates & faculty members for their inputs and valuable insights.

References

[1] SUBHASHCHANDRA TIWARI:"AUTOMATIC PET FEEDER USING ARDUINO":INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY,VOL. 7, ISSUE 3, MARCH 2018

[2] Tessema GELILA BERHAN:"PROGRAMMABLE PET FEEDER":INTERNATIONAL JOURNAL OF SCIENTIFIC ENGINEERING AND RESEARCH (IJSER),ISSN (ONLINE): 2347-3878, IMPACT FACTOR (2014): 3.05

[3]Aasavari Kank:"Automatic Pet Feeder":Pacific Academy of Higher Education & Research University

[4]Kanisius Karyono:"Smart dog feeder design using wireless communication":2016 International Conference on Computer, Control, Informatics and its Applications

Web References:

1. WWW.AUTOGATES.COM.MY
2.www.usautomatic.com
3.www.amazinggates.com
4.www.microchip.co.uk
5.www.quora.com
6.www.engineeringproject.com
7.www.nevonprojects.com
8.www.hackster.io

Smart Noise Detector System

Mr. Mohit Pandey, Mr. Shivam Pathak, Mr. Aakash Peringathara
Information Technology, Thakur College Of Engineering and Technology, Mumbai
mohit.pandey.c.610@gmail.com, shivampathak704@gmail.com, akashjperi4@gmail.com

Abstract—**IoT, i.e. Internet of Things, is currently the foremost trending field within the Branch of knowledge Communication Technology. Pollution is taken into account because the major problem that we all face in our day to day life. This is often a serious problem especially in areas where there are hospitals, schools, baby care centers, etc. With rising pollution near hospitals, schools etc., it's taken a significant toll on students, patients, etc. So a strict action should be taken and control the pollution. Although several decibel-measuring devices are applied to the environment, which indeed must be kept quiet, the decibel-meters can only detect the pollution limited to the closest area around the device. When checked in larger spaces such as a library, problems of inaccurate data will be arised. For such problems mentioned above, maybe we use more decibel-meters around different areas to an improved solution. Such problems need to be solved like averaging the measuring data, inter-connecting the cross-data between the devices with the present decibel-meters within the market. So, the proposed system implements a mobile, small-scale and internet-based sound-detective prototyping by using Internet of Things (IoT) technology to supply an improved solution mentioned above. Proposed approach is to form use of cheaper materials and low-precision sensors and single-chip microcomputer like Arduino with Bluetooth connection ability to develop an ardent sound detection device. The device will be developed under very low-cost with none display thereon, but all sensor data are going to be sent into the android application and be displayed on mobile devices.[1]**

I. INTRODUCTION

More than 72% of Indian population lives in some villages and the rest in towns and concrete agglomerations. In 1991, there were 23 metropolitan cities in India, which increased to 35 in 2001 and 53 in 2011. Thanks to this growing density, noise has gradually and steadily increased and is taken into account among the foremost pervasive and frustrating sources of everyday annoyance. Noise is taken into account as an environmental pollutant due to the danger it poses to health like Cardiovascular and Metabolic disease and also its impact on our well-being and quality of life. Noise management is a vital component within the smart cities vision, is supported by new sensors that are capable of detecting and processing various parameters associated with noise. The net of Things (IOT) enables the connectivity, accessibility and tasking of the sensors to enable the citizens to access real time information on the noise levels. This may help them in planning various activities of their daily routine in a much smarter way. Since noise incorporates a spatial component, GIS (Geographic Information System) can help to provide an appropriate and useful platform for mapping and analyzing noise sources and ultimately help for its management. It's possible to get an accurate map of the acoustic situation and Noise Risk zones (NRZ), with the assistance of advanced interpolation techniques in GIS supported by a limited number of sample points recorded using the microphone of smartphones as a sensor.

Integrating crowd sourcing techniques, it's possible to develop an efficient noise mapping system that may show the dynamically changing noise environment in a very city and thus be successful for certain authorities.[5]

II. SUBJECT HISTORICAL DATA

According to the Ministry of Environment, Forest & global climate change, energy released by any source in a very medium creates vibrations of molecules. The molecules start vibrating within the oscillatory mode and also the energy travels through the medium in kind of vibrations. If the oscillations within the medium are within the range of 20 Hz to twenty KHz it's audible by human ears and categories as sound. Ambient Noise Ambient noise is usually stated as environmental noise. It's the noise that one hears in day to day life and doesn't include workplace noise. The prime sources of noise are traffic, construction sites, industries, road-traffic, rail traffic, and structure, indoor sources such as air coolers, radio, television, air conditioners, and other home appliances. [5]

The Committee on pollution Control constituted by Central Pollution control panel, recommended noise standards for ambient air and for domestic appliances, construction equipment and automobiles as given below.

Table 1: Data of noise limits in different zonal areas

Code	Category of Area /Zone	Limits in dB	
		Day Time (6.00 am to 10.00 pm)	Night Time (10.00 pm to 6.00 am)
A	Industrial area	75	70
B	Commercial area	65	55

A. Subject Learning Outcomes and Contents

The goal of the smart noise detector is to produce a means to detect the noise and notify the needed individuals. It is of great benefit to hospitals, schools, colleges, universities residing in the silence zone. It can assist the government. Officials to require the required action against

the violations of rules in silence zones. Thus, the project is helpful in establishing authority in society.

•We present a sensible noise detector that works over IOT in order that noise is monitored.

• Our project should allow the needed officials to examine the regions violating the laws.

• The project should also display the noise levels over the android smartphones. Once a

Regions cross the brink level, it can be notified on smartphones.Areas around the 100 meters of premises such as courts,
hospitals and educational institutions are silence zones. Declaration of silence zones are done by the necessary authorities. Use of vehicular horns, bursting of crackers and
loudspeakers shall be banned in these zones. Most monitoring stations where high levels of noise were recorded were next to busy and congested roads where vehicles were the most important contributors with constant honking aggravating the matter. The ambient noise levels in residential and commercial areas indicate average day time noise levels within the metros to be more than the national standard of 55dB for the residential and 65dB for the commercial area. Because the data is obtainable on real-time, the governments can devise traffic plans for actual time intervention. [4]

III. SUBJECT DESIGN AND CURRICULUM

The design of this subject is predicated on practical hands-on knowledge of the sphere. The rationale behind this design approach is that at the successful completion of the topic, the coed should be ready to independently perform a digital forensic task. Also, the topic design caters to the requirements of two cohorts, experienced professionals working during this field and therefore the newcomers within the field. For knowledgeable professionals, the topic provides advanced topics like smartphone, web and email forensics. they will hinge on this data to further enhance their technical skills as legal skills. For the beginners, the topic starts with building the muse in digital forensics. The stress then is on the legislation, policies, procedures, technical knowledge, and preparation for reporting on digital forensic investigation cases. There's lots of emphasis on hands-on projects and solving the cases. The main focus is to develop a mindset of how evidence should be acquired and presented within the court of law that's admissible as a bit of digital evidence. For this reason, students are encouraged to create upon computing and technology knowledge, investigative skills, logic, familiarity with the local legislation and lots of passion.

• To suit the notification, the location of the violation can be provided using our project.

B. Subject Requirements

Hardware Requirements:

- **Arduino Uno (Soldered)**

The Arduino Uno is an open-source microcontroller board. It is developed by Arduino.cc. The board is provided with various sets of digital and analog input/output (I/O) pins that will be interfaced to expansion boards (shields) and numerous other circuits .The board has 6 Analog pins, 14
Digital pins and programmable with Arduino IDE (Integrated Development Environment) with USB cable. It is powered by the USB cable or by an external 9-volt battery, though it accepts voltages between 7 and 20 volts. It's also like the Arduino Nano and Leonardo. Commons Attribution Share- Alike 2.5 license offers the imaginative hardware reference design on Arduino website. Layout and production files for a few versions of the hardware are available.

- **Sound Sensor Module**

Generally, this module is employed to detect the intensity of sound. The applications of this module mainly include switch, security, still as monitoring. The accuracy of this sensor is changed for the convenience of usage. This sensor employs a microphone to produce input to the buffer, peak detector and an amplifier. This sensor notices a sound, & processes an output signal to a microcontroller. After that, it executes required processing. The sensor has the ability to detect the noise level decibels within the range of 3 kHz - 6 kHz frequency. In smartphones, there's an android application namely decibel meter accustomed to measure the sound level.

- **HC-05 Bluetooth Module**

HC-05 may be a Bluetooth module which is meant for wireless communication. This module is employed in a master or slave configuration. *Smart Noise Detector System*
Bluetooth serial modules allow all serial enabled devices to speak with one another using Bluetooth.

HC-05 Bluetooth module can communicate with smartphones using a Bluetooth terminal application for sending and receiving data. Various Bluetooth terminal applications for android and windows in respective applications can be found.

- **Android Smartphone**

The android is a robust operating system and it supports a huge number of applications and softwares in Smartphones. These applications are more attractive, advanced and comfortable for the users. The supporting hardware is based on ARM architecture. As android is an open source operating system, it can be used for free by anyone and can easily be developed further.

C. Assessment Items

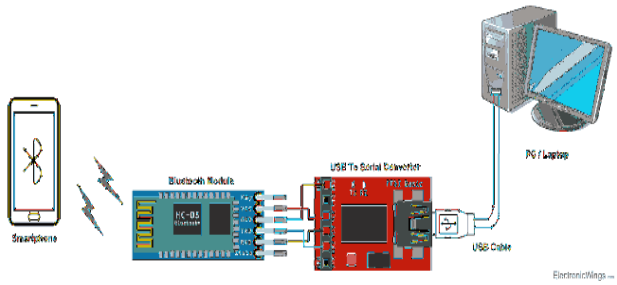


Fig 2: Data sharing

The neighborhoods of hospitals are declared as silence zones in order that patients receiving treatment don't seem to be disturbed. So are areas around schools, colleges, courts and other such places that are required to be noise-free.

According to the standards, the decibel level in silence zones mustn't exceed 50 dB during the day and 40 dB during the night. The noise monitoring network can be further expanded to provide informative noise maps. Using such maps, we can install various barriers such as enforcement of speed limits for heavy vehicles traversing through residential areas, demarcation of no-honking zones and adoption of proper land use plans in trouble spots like hospitals, libraries and various sensitive buildings.

This project was developed to produce a convenient method of detecting noise in silence zones. This project, as far as feasibility is worried, didn't cost an enormous amount because we used Arduino Uno .This project is implemented around a value of Rs.800 for every setup .The materials used are easily available and thus the project is extremely feasible and may be expanded as required.

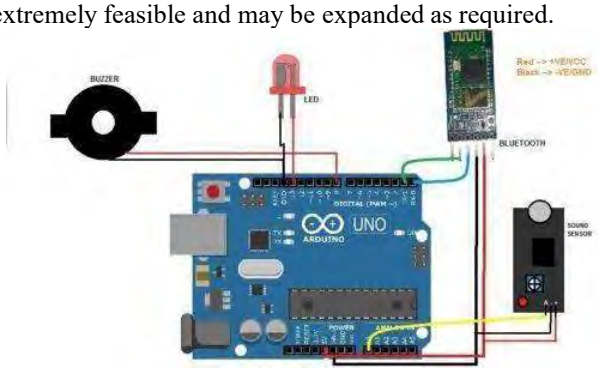


Fig 3: circuit Diagram of system

D. Final Examination

Our system detects the noise within the silence zone using the Sound Sensor. Sound sensor detects noise in decibels. The default intensity is entered within the program in a similar way of decibels. The system is placed in a very very silence zone and when the amplitude crosses the sting level, it's notified on a buzzer and a LED for a certain period of it slow. The system also uses a Bluetooth module to share data to an android device.

Data is shared over Bluetooth to an android application. The android application provides us the selection to modify the threshold and thereby control the amplitude in a very very surrounding. The GPS module is utilized to provide the location of the violation.

Thus, we are able to detect noise successfully in a very very silence zone and control the threshold level using an android application.

For Testing, power the Arduino and open the app tap at the Bluetooth Icon and choose HC 05 from the list to connect your phone to Bluetooth. Now you will see the amplitude value within the text bar. When amplitude reaches the sting value, the LED light and Buzzer will automatically get triggered by Arduino. User can also set a threshold value using the app. To do so, tap on the text box next to the 'send' button and enter the price you'd wish to feature and click on on the 'send' button to line the sting value.

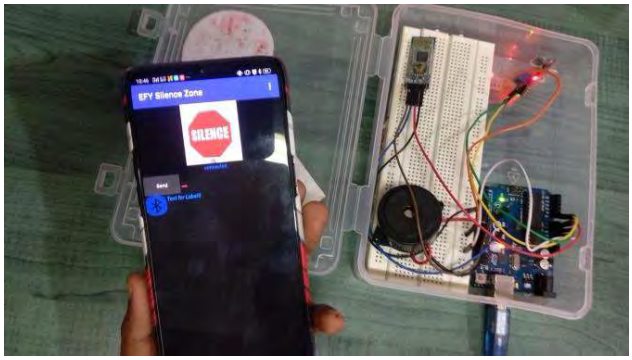


Fig 4: Final Kit of system

IV. CONCLUSION

This project mainly consisted of detection of noise in silence zones. Declaration of silence zones are done by the necessary authorities. Use of vehicular horns, bursting of crackers and loudspeakers shall be banned in these zones. Most monitoring stations where high levels of noise were recorded were next to busy and congested roads where vehicles were the biggest contributors with constant honking aggravating the matter. The ambient noise levels in residential and commercial areas indicate average day time noise levels within the metros to be above the national standard of 55dB for the residential and 65dB for the commercial area.

This project presented a framework using which attendance management is created automated and on-line. Fingerprint Identification System used for student identification is quicker in implementation than the opposite fingerprint identification systems. Biometric

NOISE DETECTOR gives a wonderful assistance to high school management because it complies with authenticated student attendance requirements of universities authorities & enhance the status of a college to face with the only within the peer set as a up to date, progressive and iconic institution giving students, teachers and administrators the way of pride and positivity. For straightforward and easy understanding basic SQL is utilized .The proposed system saves college precious time and value, session on session which could be instead invested in improving the academic systems, student development practices and developing a desirable college culture. It brings in simplicity, consistency and accuracy in College operations and reports which suggests lesser hassles in reconciliation, mistake and error management.

In the future, this project is extended to convey more detailed analysis and solve problems like attendance, maintaining manual record of information, studying the data manually and will help to avoid wasting plenty of time. Also, it's accustomed to predict the outcomes and suggest measures which can be taken to forestall an error and significantly it'll help to chop back labor costs. Regarding our fingerprint identification system, we are planning to introduce more indexing techniques like ridge density tolerance etc. for creating the search faster. Also the keys used are more efficient when complex keys are visiting. We try to reduce matching error rates. The project relies on a static calendar and in future we are determined to create it flexible for dynamic. We are able to show graphical representation of the attendance which may be very easy to interpret.

References:

1. Yung Chung Tsao1, Bo-Rui Su2, Chin-Tan Lee3, Chia- Chun Wu4,"An implementation of a distributed sound sensing system to visualize the noise pollution",2017 International Conference on Applied System Innovation (ICASI).
1. Prajakta S. Palkar1, "Microcontroller based Noise Detection and Controlling Electronic Appliances", 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2019.
2. Palaghat Yaswanth Sai1, "An IoT Based Automated Noise and Air Pollution Monitoring System", International Journal of Advanced Research in Computer and Communication Engineering, 2017.
3. Arushi Singh1 , Divya Pathak2 , Prachi Pandit3 , Shruti Patil4 , Prof. Priti C. Golar5 , "IOT based Air and Sound Pollution Monitoring System", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,2017
4. Pankaj Tarone1 , Kunal T. Nikam2 , Anirban Bhowmick3 , Yogesh Shinde4 , Surya S.Durbha5, "Urban Noise Detection and Management for a Smart City using IOT Enabled Model ", 17 there India User Conference 2017.
5. Dennis Menezes1 , Nachiket Waikos2 , Zameer Khazi3 , Manish Jha4 , Kanchan V. Wankhade5, "IOT based Air and Sound Pollution Monitoring System", International. Journal for Research in Applied Science & Engineering Technology (IJRASET), 2017.

Voice Controlled Assistant Using Raspberry-Pi

Yash Bhatt, Hetvi Shah, Tanmay Phalke
Tcet, Mumbai
bhattyash031@gmail.com, hetvishah208@gmail.com, tanyphalke1210@gmail.com

Abstract—**As technology is entering a new era of computing that many are calling the Internet of Things (IOT). System to System, System to infrastructure, System to Environment, the Internet of Everything, the Internet of Intelligent Things, Intelligent systems—call it what you want, but it's happening, and its potential is huge. A voice controlled Assistant offers voice commands, voice searching, and voice-activated device control, letting you complete a number of tasks after you've said the "OK Google" or "Hey, Google" wake words. It is developed to give you chatty communications. Here we propose an IOT based Voice Controlled Assistant project using Raspberry Pi3. This Raspberry Pi Voice Controlled assistant project will walk you through your very own Pi-powered voice controlled Assistant. This assistant will actively listen to your commands and respond to your queries, all we need to say is "Ok Google" or Hey Google‖ followed by your query. The System consists of the Raspberry Pi3 Board with Speaker and Mic that are connected to it. The system starts just by saying "OK GOOGLE", followed by the query. The System records the query from the Mic, processes the query and the reply is given from the speaker that is connected to the raspberry pi board.**

Keywords –, voice searching, voice-activated device control , chatty communications, Ok Google.

I. INTRODUCTION

"IF YOU PONDER THAT THE INTERNET HAS CHANGED YOUR LIFE, THINK AGAIN.

The Internet of Things is about to transform it all over again!"

A Voice Controlled Assistant focuses on voice commands, voice searching, and voice- activated device control, letting you complete a number of tasks after you've said the "OK Google" or "Hey Google" hot words. It is designed to give you interactive interactions. Our project Google Assistant Using Raspberry Pi3 mainly focuses on conversation between the human and the

Assistant to get your query solved or to get the information about any smallest thing happening around.

Voice Controlled Assistant Is mainly designed to reduce the human efforts. Instead Of Entering the Browser and typing the query and search for it takes a lot of time. Voice Controlled Assistant Makes It simpler for us. All we need to do is give a command "OK! Google" and ask the query and the assistant will give the desired results within seconds.

All we need to do is setup the whole assistant on raspberry pi. After successful setup of the assistant we need to run various scripts, programs and commands. After successfully completing it Our Pi-powered assistant is ready. All we need to do is say "Ok Google" to make it active followed by your questions like ―How are you or tell the traffic update It will give the answers quickly.

II. PROBLEM DEFINITION

Design a smart IOT assistant that helps us automating things that we do in our day to day life that eliminates input via text and takes ―voice as input which helps us getting things done on the go. To build a system which provides you appropriate knowledge when asked for. To develop an assistant which performs various tasks for you such as playing songs, giving you the traffic update ,telling you the weather or like making a shopping cart for you or playing a joke. It also can interact with the user in various languages. This device has lesser cost than the original GOOGLE HOME and has all of the features of it. It works in two languages English as well as Hindi.

III. BACKGROUND

Voice Controlled Assistant is an Artificial Intelligence (AI) based virtual assistant developed by Google that is mainly available on mobile and smart home devices. Unlike the company's previous Virtual

Assistant. Google Now, the Voice Controlled Assistant can engage in two-way conversations.

Voice Controlled Assistant initially debuted in May 2016 as part of Google's messaging app Allo, and its voice-activated speaker Google Home. After a period of exclusiveness on the Pixel and Pixel XL smartphones, it began to be positioned on other Android devices in February 2017, counting third-party smartphones and Android Wear (now Wear OS), and was introduced as a unique app on the IOS operating platform in May 2017. along side the announcement of a software development kit in April 2017, the Assistant has been, and is being, further stretched to upkeep a large diversity of devices, including cars and third party smart home machines. The functionality of the Voice Controlled Assistant can also be boosted by third- party developers.

Now the Voice Controlled assistant can be integrated with any of the devices or the system. We also have integrated the Google assistant in our project using the raspberry pi.

IV. IMPORTANCE

Google Assistant is focused on reducing human efforts. Google Assistants are widely used on mobile phones, speakers, Television, Cars etc. Our main intention was to develop a Google Assistant On the different system which makes use Of Internet Of Things. We thus developed a Google Assistant on the Raspberry PI using various components. Importance of ―Google Assistant Using Raspberry Pi3‖ are as follows:-

- Create new calendar events on the fly by telling the Assistant "add to my calendar" or "make an appointment" followed by the event's name, date, time, and optionally even location.
- Get a broader overview of your day by activating Assistant and saying "tell me about my day" or even just "good morning."
- Asking Google Assistant to Play a song or even tell you jokes.
- Getting the traffic or the weather updates by just asking a question.

V. OBJECTIVES AND SCOPE

The main objective of developing this project was to build a Pi-powered Voice Controlled Assistant which interacts with people in human language. The user just needs to activate the system by saying ―OK! GOOGLE followed by the query say how are you? And the assistant will take the input and give the output by giving the

Appropriate, answer. Various Objectives fulfilled by the project are:-

User can create new calendar events on the fly by telling the Assistant "add to my calendar" or "make an appointment" followed by the happening's title, date, time, and voluntarily even location. User can get a broader overview of the day by activating Assistant and saying "tell me about my day" or even just "good morning.

User can ask Voice Controlled Assistant to Play a song or even tell you jokes. User can get the traffic or the weather updates by just asking a question. In the future this assistant can be used to develop a Intruder Detection System using Fingerprint Sensor by capturing the fingerprints and allowing only authorized users to enter the system and sending alerts if intruder tries to enter the system. It can also be used For Home Automation which is trending topic in the real world.

VI. PHASES

Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: Getting strong knowledge of the project title and doing exploration on it, we will get our description. Preparation: After scrutiny we will first study about it and do some research on the project for our better understanding and also getting a rough depiction about what would be our problem definition for this particular project.

Designing: Then we will build the design of the project and agreeing to that, will list down all the necessities needed for the creation for the prototype of our project. Implementation: After gaining the requirements, we first developed the connections of the hardware components (Raspberry Pi 3, Microphone, Speaker) and then added the coding part.

Phase 2: Testing and Deployment.

Testing: After the model is ready we will first associate the hardware with the allotted code and then we will check if it supports the machine or not. If not we will resolve the issues relating to it and will crisscross again.

Deployment: After completion of integration and testing of project, real time testing and operation of the system will be done.

VII. METHODOLOGY

This proposed method takes an automatic control action upon detection of hot word ―OK GOOGLE!. The assistant will get invoked and start running its function to read the command spoken after detection of the hot word. It then processes the input and then gives the desired output after connecting with the Google servers. Then it gives us back the processed output through the connected Speaker.

VIII. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware

- 1 Raspberry Pi 3
- 1 Speaker
- 1 Microphone
- 1 HDMI to VGA converter
- 1 USB cable

Software

- NOOBS OS
- PYTHON 3
- LINUX
- Google SDK

IX. BLOCK DIAGRAM

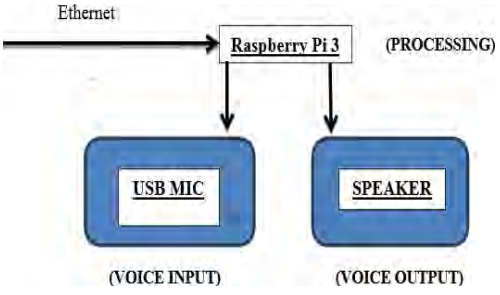


Fig. Block Diagram

The above block diagram represents the actual working of our project. Here the voice input is given to the mic which is processed using the raspberry pi and the reply is given through the speaker in human language.

X. CIRCUIT DIAGRAM

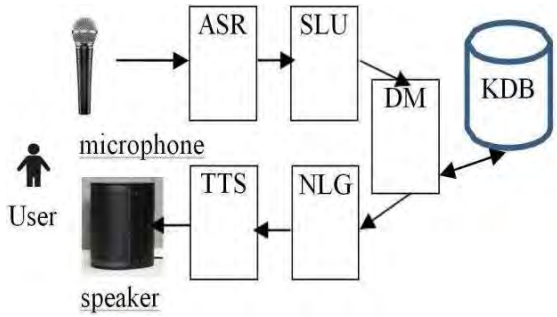


Fig. Circuit Diagram

This circuit diagram represents the internal working of our assistant. It consists of (ASR) Automatic Speech Recognition Part, (SLU) Spoken Language Understanding, Dialog Manager (DM), A Knowledge Database (KDB), A Natural Language Generator (NLG) and a Text To Speech Synthesis (TTS) system which makes the Google assistant fully functional.

XI. OUTPUT

When we say OK Google! The assistant starts running and reads input in the form of voice and does the task according to the input received.

XII. RESULT

When we say OK Google! The assistant starts running and reads input in the form of voice and does the task according to the input received.

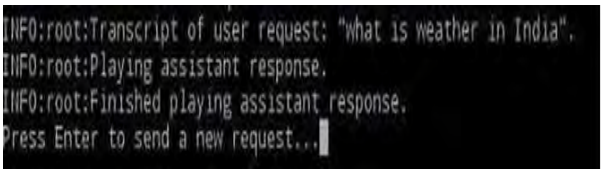


Fig. Voice Controlled Assistant working Kit

XIII. OUTCOMES

Creating the outline, works survey and achievability study for proposed solution which will act as an input to the design phase.

Suggesting Statement of work, scope description and scope border for planning the model from the problem definition to decide what needs to be done and what not to be done.

Phase 1 Implementation:

Implementation of the Raspberry PI model with the help commands running it in the Linux environment.

Phase 2 Testing:

Testing was done with cases like testing mic and speakers , Testing by asking different queries to the assistant .The testing of the model was a success are mentioned in the above figures.

Phase 2 Deployments:

Real time working was done when we asked questions and assistant replied the answers appropriately. Thus the model works efficiently.

XIV. RESEARCH GAPS

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. They are as follows:

- Since the Wi-Fi was not working with the Raspberry Pi, the device cannot be connected to the home Wi-Fi right now.
- API such as YouTube and Google Play Music have not been installed currently due to errors in Google SDK.
- The device cannot control other devices connected on the network right now. It needs to be configured in that manner to do so.

XV. SUMMARY

A Voice Controlled Assistant focuses on voice commands, voice searching, and voice-activated device control, letting you complete a number of tasks after you've said the "OK Google" or "Hey Google" keywords. The project focuses on building an IOT based ―Google

Assistant‖ using Raspberry PI3 platform. Before the project starts there is the abstract and the introduction of our project which gives the brief idea about the project. It also consists of the Importance of the project, Objectives and the future scope of the project. Thus, this project was a success with all the use cases and fulfilling all the phases of the project development anddeployment.

XVI. CONCLUSION

We have successfully shown that the implementation of an Assistant by IOT based project is possible. It fulfills the various commands commanded by the user. The Assistant takes the input from the user through the Microphone connected. It gets activated when the user gives the input by saying “OK Google” and according to the given input command, the output is obtained to the user. The weaknesses of the system are that currently it is not able to perform the Google Play API and also does not support Head Listyet.

References

1. https://www.researchgate.net/publication/329817651_Implementation_of_Google_Assistant.
2. https://www.techrepublic.com/article/google-assistant-the- smart-persons-guide/
3. https://www.c-sharpcorner.com/article/diy-google-ai-assistant-using-a- aspberry-pi2/
4. https://www.makeuseof.com/tag/diy-google-home-assistant-raspberry-pi/
5. https://pimylifeup.com/raspberry-pi-google-assistant/
6. http://www.iosrjournals.org/iosr-jece/papers/Vol.%2014%20Issue%203/Series-1/G1403013639.pdf
7. https://www.hackster.io/Salmanfarisvp/googlepi-google-assistant-on-raspberry-pi-9f3677

Saline Monitoring System Using Ardiuno

Pratimesh Majumdar, Shivam Kumbhar, Pratik Kumbhar
TCET, Mumbai
pmajumdar912@gmail.com, shivamkumbhar432@gmail.com, pratikkumbhar121999@gmail.com

Abstract— **IV saline is used to replenish lost fluids, flush wounds, deliver medications, and also in surgery. A constant monitoring of the saline level in the bottle is required because if the saline in the bottle is fully consumed and the bottle is not replaced immediately then the pressure difference between the patient's blood pressure and the empty saline bottle causes an outward rush of blood into the saline. This situation can pose a serious threat to the patient's well-being. In current health care measures whenever a saline is fed to any patient, they need to be continuously administered by a nurse or any caretaker.**

In this paper we have proposed an Internet of Things (IOT) based saline monitoring system. The project is made on Arduino platform and uses infrared sensor to indicate if the saline level falls below a certain level. The working of this model is based on the phenomenon of refraction. When infrared rays pass through different medium refraction occurs which causes a reduce in the voltage received at the receiver. This system notifies the concerned authorities, if such is the case, which is achieved by triggering the buzzer.

It is meant to improve the quality of health-care provided at a time when the work-load on hospitals is ever increasing. Having one less concern allows them to concentrate on different tasks which not only proves beneficial for the patients, but also for the staff. If the saline level falls the staff is alerted, who can then prioritize the timely change of saline IV, thus preventing the above-mentioned concerns as well as lead to the effective utilization of their time. The solution provided can be made at a relatively low cost.

Keywords – Arduino; Saline IV; Hospital; IOT

I. INTRODUCTION

The working condition in hospitals may lead to improper health care monitoring. This might be due to excessive number of patients or the inattentiveness of relatives. If, in such a case, the saline bottle gets empty, the pressure difference between the patient's blood and the empty saline bottle causes a vacuum to be created which causes the blood to be sucked out from the patient's body causing excruciating pain. The only way to solve this is to change the saline IV bottle. In such cases prevention is the only feasible solution. The Saline Monitoring System constantly monitors the saline bottle and helps to detect whether the saline level has fallen below a certain level. This results in the reduction of hemoglobin level of patients and may also lead to shortage of red blood cells (RBCs) in the patient's blood causing tiredness. Therefore, there is a need of developing a saline level monitoring system which will reduce the patient's dependency on the nurses or caretakers to some extent. If it does, the system alerts the concerned work-staff. Thus, they are reminded of this only when their concern is necessary leading better work flow as well as providing a fast and efficient response.

The reverse blood flow might at times be so intense that it infuses into the saline bottle and colors the whole fluid red. Though there is little or no harm to the patient, the increased pressure causes intense pain, alerting and diverting the attention of the attendants and the nursing staff. Thus, it simply induces uneasiness in the patients and increases the work of the staff as they need to clean that up. The proposed system can not only be used for monitoring saline level in hospitals but also can have many different applications.

With the increasing demands in the healthcare sector the quality of care provided has been falling. With the increase in populace and the deterioration of their lifestyle, the number visits to the hospital has considerably increased in the present time.With such high demand many hospitals cannot keep up with their quality of healthcare. Our project is a small step in improving the care provided by such hospitals to the patients as well as reduces the stress of the staff and nurses.

The goal is to design a microcontroller (IoT) based Saline Monitoring System that will continuously monitor to check whether the saline level has fallen below the required minimum level and if yes, activate the alarm to notify the staff.

II. EXISTING APPROACHES

In the current healthcare measures, professional nurses are responsible for managing, monitoring and providing care to patient receiving saline. Roller clamp is used for manually controlling the saline infusion rate at the hospitals. If roller clamp rolls in one way, it compresses the intravenous tube more tightly which make tube more thin and allow saline fluid to flow through at a slower rate.

In the present world, there is no such monitoring system which will reduce the dependency of the patients on the nurses, doctors and would also reduce the need for the nurses to go to patient's bed every time to check saline level status of each patient. Therefore, there is need for development of IoT based saline level monitoring system

III. IMPORTANCE

- The requirement of the saline level detection and its alert system is because of the increase in the negligence of the hospital staff caused by factors such as increased work load, lack of sleep due to overworking or simply forgetting about it.
- Hence, this project is of great importance to manage the staff's time, leading to helping both the patient and the staff. Importance of the Saline Level Detection and its alert system are as below:
- Alert system is primarily designed to alert the staff of a drop in the saline level so they can safely change the saline bottle.

- Correctly maintaining and operating the alert system is effective in improving the work flow.
- Better health care can be provided by giving timely alerts.

IV. FEATURES

The Project consists of the following features:

- Alerting mechanism: In order to alert the staff of the low saline level a buzzer is used.
- The buzzer will ring if the saline level falls below a certain level.
- High sensitivity: The Infrared sensor used is sensitive to the amount of saline present at each point of time in the bottle. As soon as it senses that the saline is below the threshold level it triggers the alerting mechanism.
- Cost effective: The overall cost of the project is low and can thus be easily implemented.
- Simple circuitry: The connections are simple and hence it is easy to develop.

V. PROPOSED METHODOLODY

The proposed method takes automatic action upon detection of saline level through constant measuring of saline level. Initially while the saline bottle is full the voltage at the receiver of the phototransistor is low. This is due to the diffraction of the IR rays through the solution. When the solution falls below the level at which the IR phototransistor is placed the voltage increases as there is no diffraction. This sends a high signal to the microcontroller which triggers the alarm and notifies the ones concerned.

VI. SYSTEM REQIOREMENTS

The Arduino UNO is developed by Arduino.cc and is an open-source microcontroller board based on the Microchip ATmega328P microcontroller. The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits. It has 14 Digital pins, 6 Analog pins, and is programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable. IR phototransistor and receiver Infrared phototransistor is a special dark lens that reduces sensitivity to visible light. It can only trigger on with infrared light.

Infrared Phototransistor

An Infrared light emitting diode (IR LED) is a special purpose LED emitting infrared rays ranging from 700 nm to 1 mm wavelength. An IR LED is a type of diode or simple semiconductor. Electric current is allowed to flow in only one direction in diodes. As the current flows, electrons fall from one part of the diode into holes on another part. In order to fall into these holes, the electrons must shed energy in the form of photons, which produce light. An IR sensor is a device that detects IR radiation falling on it. Buzzer A buzzer or beeper is an audio signaling device which may be mechanical.

In common parlance a Buzzer is a signaling device that is not a loudspeaker. It can be mechanical, electromechanical, or electronic (a piezo transducer).If the saline level is below the preset threshold the buzzer would start buzzing and alert the nurse or the doctor.

VII. ARCHITECTURE

Fig.1: Block Diagram

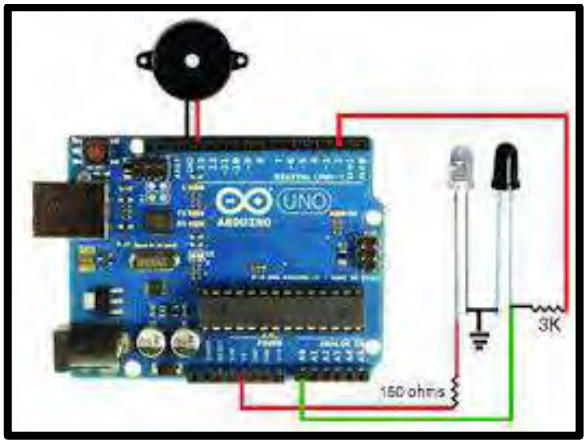


Fig.2: Circuit Diagram

VIII. WORKING

Initially while the saline bottle is full the voltage at the receiver of the phototransistor is low. This is due to the diffraction of the IR rays through the solution. When the solution falls below the level at which the IR phototransistor is placed the voltage increases as there is no diffraction. This sends a high signal to the microcontroller which triggers the alarm and notifies the ones concerned.

When the saline level is above the level at which the IR phototransistors and receivers are placed the voltage is always below 4V which shows that the rays are getting diffracted and thus the buzzer does not get triggered. when saline bottle is full Whenever the saline level falls below the threshold it is detected by the sensor which gives an output around 5V which triggers the buzzer

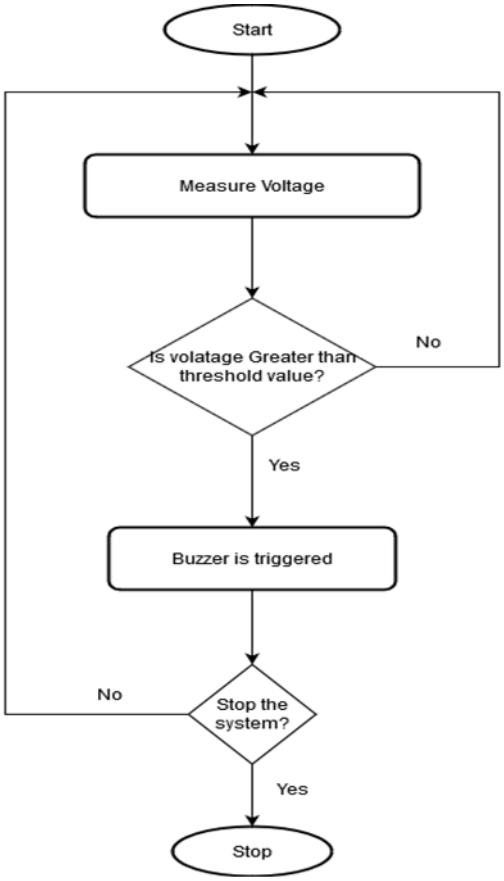


Fig.3: Flowchart of the proposed system

IX. RESULTS

When the saline level is above the level at which the IR phototransistors and receivers are placed the voltage is always below 4V which shows that the rays are getting diffracted and thus the buzzer does not get triggered.

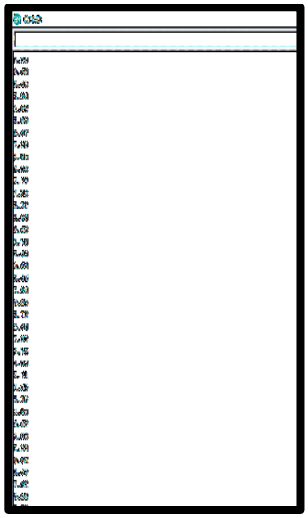


Fig 4: Voltage when saline bottle is full

Whenever the saline level falls below the threshold it is detected by the sensor which gives an output around 5V which triggers the buzzer.

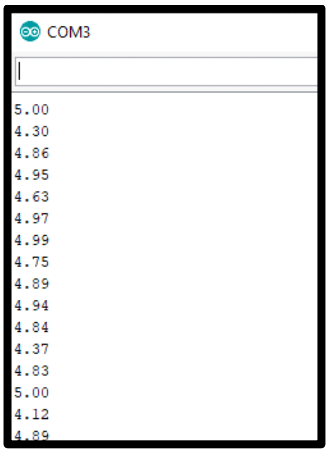


Fig 5: Voltage when saline bottle is empty

X. CONCLUSTION

We have successfully shown that the saline level monitoring system based on IOT is possible and can detect saline level, warn consumer when the saline falls below a certain threshold level. The system can constantly keep track of the saline level ensuring that the alarm is sounded reliably. This system uses the buzzer as a warning device in the event of low saline level. It increases the work efficiency of the staff as well as improves the health care provided.

References

1) Khushboo Vaishnav, Neha Swamy, Nargees Bano Haidarali, Prof. Madhuri Patil, "IoT Based Saline Level Monitoring System", International Journal of Innovation & Advancement in Computer Science, Volume 6, Issue 10, October 2017, ISSN 2347 – 8616

2) Pooja Pandit Landge, Siddharudha .S. Shirgan,"Smart Saline Level Monitoring and Control System" International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 7 Issue VIII, Aug 2019, ISSN: 2321-9653

3) Anusha Jagannathachari, ArchanaRajan Nair, "Saline level Indicator", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p-ISSN: 2278-8727

4) Pooja Kanase, Sneha Gaikwad "Smart Hospitals Using Internet of Things(IoT)", International Research Journal of Engineering and Technology (IRJET), Volume 3, Issue 3, March 2016, e-ISSN: 2395 -0056, p-ISSN: 2395-0072

Garbage Monitoring System

Anupama Singh, Dhruva Trivedi, Anushka Singh
Information Technology, Thakur College Of Engineering, Mumbai,India
anupama.singh2328@gmail.com, trivedidhruva981@gmail.com, anushkasingh1055@gmail.com

Abstract—**Garbage may consist of the unwanted material leftover from City, Public area, Society, College, home, etc. This project is related to the "Smart City" and based on the "Internet of Things" (IoT). So, for a smart lifestyle, cleanliness is needed, and cleanliness begins with Garbage Bin. Garbage Monitoring System project helps to eradicate or minimize the garbage disposal problem. Many times, in our city we see that the garbage bins or dustbins placed at public places are overloaded. It creates unhygienic conditions for people as well as ugliness to that place leaving a bad smell. To avoid all such situations, we are getting to implement a project called IoT Based Smart Garbage and Waste Collection Bins. These dustbins are interfaced with a microcontroller-based system having Ultrasonic wireless systems alongside central system showing level and current status of garbage, on an internet browser with HTML page by Wi-Fi and Email will forward Gmail account. Hence the status will be updated on the web page. A major part of our project depends upon the working of the Wi-Fi module; essential for its implementation. The main aim of this project is the enhancement of a smart city vision and reduction of human resources. This proposed system assures the gathering of garbage and waste soon when the waste level reaches its intensity. The system will thus provide correct information, increase the efficiency of the system. This proposed system is more efficient and time-saving as compared to the existing system.**

Keywords –IoT- Internet of things, IR -Infrared sensor, GSM-Global System of Mobile Communications module, etc.

I. INTRODUCTION

Internet and its applications became an integral part of today's human lifestyle. It has become an important tool in every aspect. Due to the tremendous demand and necessity, researchers went beyond connecting just computers into the online. These researches led to the birth of a sensational gizmo, Internet of Things (IoT). Communication over the web has grown from user-user interaction to device – device interactions lately. The Sensor Network Lab concepts were proposed years back but still, it's in the initial stage of commercial deployment. The home automation industry and transportation industries are seeing rapid growth with Sensor Network Lab. Yet not many articles are published during this field of study. This paper aims to structure a state of the art review on Sensor Network Lab. The technology, history, and applications are discussed briefly alongside various statistics. Since most of the process is done through the internet we must have an active highspeed internet connection. The technology is often simply explained as a connection between humans-computers-things. All the equipment we use in our day to day life can be controlled and monitored using the Sensor Network Lab. A majority of the process is done with the help of sensors in Network Lab. Sensors are deployed everywhere and these sensors convert raw physical data into digital signals and transmit them to its centre. In this way, we can monitor environmental changes remotely from any part of the world via the internet. This systems architecture would be based on the context of operations and processes in real-time scenarios. Smart collection bin works similarly with the combination of sensors namely weight sensor and IR sensor that indicates its weight and different levels respectively. The IR sensors will show us the varied levels of garbage within the dustbins and also the load sensor gets activated to send its output ahead when its threshold level is crossed. These details are further given of the microcontroller (ARM LPC2148) and the controller gives the details to the transmitter module (Wi-Fi module). At the receiver section, a mobile handset is required to be connected to the Wi-Fi router, therefore, the details of the rubbish bin are displayed onto the HTML page in the web browser of our mobile handset.

II. BACKGROUND.

Learn cleanliness from the cat is claimed and appreciated in Ethiopian society. The scenario beyond this proverb is that keeping the environment clean is quite important and it is used to emphasize more on giving attention to the cleanliness. Nonetheless, it's not as easy to speak out the proverb to effectively and properly managing the garbage. We frequently observe garbage bins being filled over and extra waste materials being disposed and accumulated around the bin in several cities of Ethiopia. That improperly disposed of garbage will be the dwelling for a various number of dangerous microorganisms, insects, and mosquitoes to breed on. Because of this, the severe and contagiousdisease is stimulated and also bad smell comes out of it and may cause illness to human beings. The municipality of most cities in Ethiopia has strived its best to alleviate this problem by providing several, garbage bins throughout the town. However, it is a manual approach and several trucks from the municipal authority is sent to the waste bins to collect the waste. The wastes are loaded to the truck and conveyed to the pre-specified locations. Because of this, the category of the people involved in collecting and transporting the wastes are usually not responsible enough to do the job well done. Very often the wastes are not collected from every waste bin properly due to municipal authorities did not have information about the garbage bin. The manual waste collection and management approach has problems such as a lack of information about the collecting time and place. Because of this, it is and less effective i.e. trucks go and may get empty garbage bin. Generally, there is a lack of a proper monitoring system to follow all activities related to waste management and lack of smart monitoring of the condition of the bin. This proposed paper shows an effective solution to manage, garbage. This garbage monitor is implemented using sensors, and

Raspberry Pi microcontroller. The details of each dustbin are monitored by the municipal authorities with the help of GUI. The implementation of a garbage monitoring system using sensors, microcontrollers and GSM modules assures the cleaning of dustbins soon when the garbage level reaches its maximum and also notify the people not drop the dust outside the dustbin when they try to drop it out of the bin by sensing their motion. If the dustbin is not cleaned in a specific time, then the record is sent to the higher authorized office who can take appropriate action against the concerned contractor [2]. This system also helps to watch the fake reports and hence can reduce corruption within the overall management system. This reduces the total number of trips of Automated Garbage Monitoring System Using Raspberry Pi.

A. SYSTEM ARCHITECTURE.

Fig. 1 describes the diagram of the monitoring system. It consists of several ultrasonic sensors,
the Raspberry Pi Uno microcontroller, and the GSM module. The ultrasonic sensors are wont to detect the extent of garbage in each bin, and it'll send this information to the Raspberry Pi Uno, which acts as the system controller. In the case where the bins are
already full or almost full, then it'll generate a warning message which can be sent to the municipality via SMS, by using the GSM module. Moreover, this
system is also connected to several LEDs, in green or red colors, in order to alert all the residents regarding the bin status.

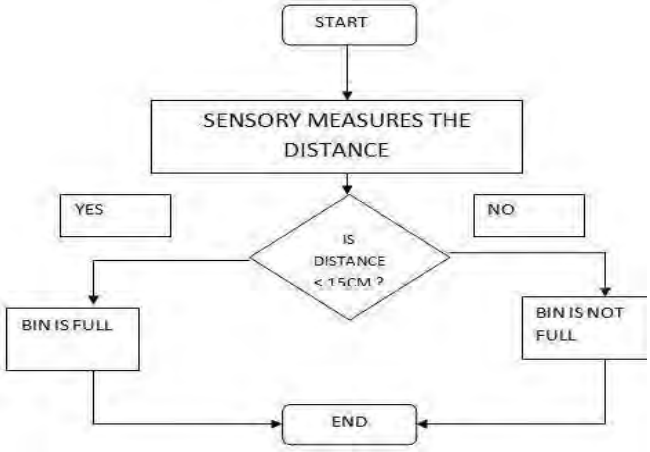


Fig. 1- Methodology.

B. SYSTEM PROCESS FLOW.

At first, the level or the height of the garbage in each bin is measured by using the ultrasonic sensor. This information is then received and processed by the Raspberry Pi Uno. It will determine whether the garbage level has been surpassing the threshold level or not. For this research purpose, there are two threshold levels sets: the first threshold is at 70% of the bin height, and the second threshold is set at 90% of the bin height. If the garbage level in the bin is crossing the first threshold level, then the first warning message is generated and sent to the municipality. Besides, the green LEDs will be turned ON in order to alert all the residents on every floor. Next, if the garbage level in the bin is crossing the second threshold level, then the second warning message is generated and sent to the municipality. In this case, all the residents will be alerted when the red LEDs are turned ON.

C. FEASIBILTY STUDY

1. Technical Feasibility - this assessment focuses on the technical resources available to the organization. It helps organizations determine whether the technical resources meet
capacity and whether the technical team is capable of converting the ideas into working systems. It also involves the evaluation of the hardware, software, and other technical requirements of the proposed system.
2. Economic Feasibility - this assessment typically involves a cost/ benefits analysis of the project, helping organizations determine the viability, cost, and benefits related to a project
before financial resources are allocated. It also is an independent project assessment and enhances project credibility— helping decision-makers determine the positive economic benefits to the organization that the proposed project will provide.
3. Legal Feasibility - this assessment investigates whether any aspect of the proposed project conflicts with legal requirements like zoning laws, data protection acts, or social media laws.
during a specific location. A feasibility study might reveal the organization's ideal location isn't zoned for that sort of business. That organization has just saved considerable time and energy by learning that their project wasn't feasible right from the start.
4. Operational Feasibility - this assessment involves undertaking a study to research and determine whether—and how well—the organization's needs are often met by completing the project. Operational feasibility studies also analyze how a project plan satisfies the wants identified within the requirements analysis phase of system development.
5. Scheduling Feasibility - this assessment is that the most vital for project success; in any case, a project will fail if not completed on time. In scheduling feasibility, a corporation estimates what proportion time the project will fancy complete.

D. SYSTEM DESIGN REQUIREMENT.

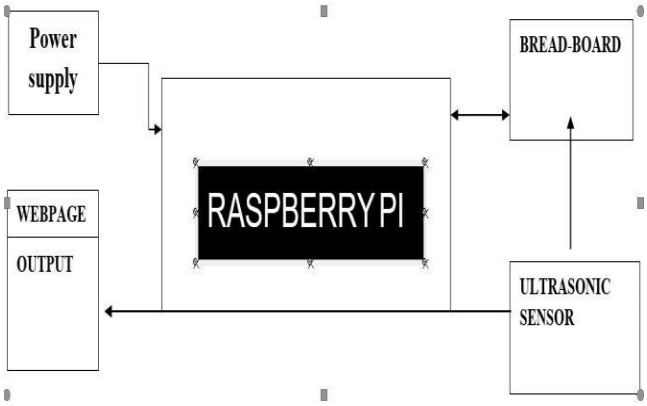


Fig. 2 – Block Diagram.

The power supply is given to the raspberry pi it activates the Raspberry pi. Then make the necessary connections with the jumper wires with the bread board and Raspberry Pi. Connect the ultrasonic sensor with the bread board. When the code runs, the ultrasonic sensor takes the input and the output is displayed on the webpage wirelessly.

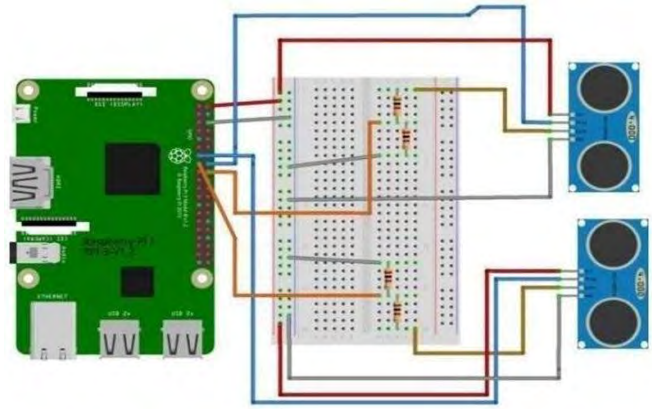


Fig. 3 – Circuit Diagram.

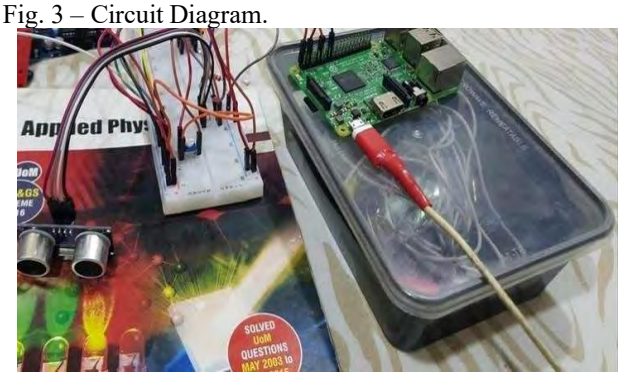


Fig. 4 – GUI Design.

In the above fig, the jumper wires are used to connect the bread-board with the GPIO pins of Raspberry Pi, and resistors are used for proper power supply for the Ultrasonic sensor, which measures the distance.

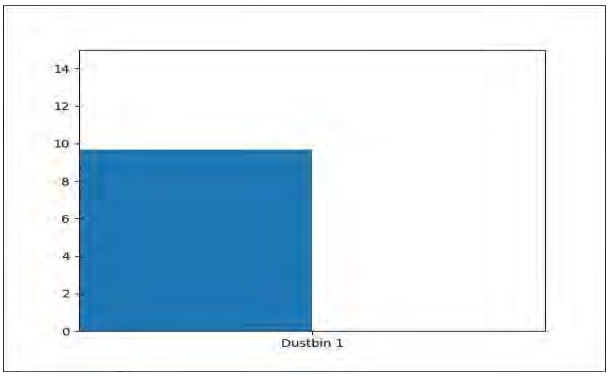
When an object is placed in front of the ultrasonic sensor it takes the input as the distance and displays whether the bin is full or not on the webpage on the laptop.

III. SYSTEM OUTPUT.

A. Outputs (Output of the project only)

Successfully implemented the garbage monitoring system. The Ultrasonic Sensor detects the level of the trash in the bins and alerts the system, of the increase in the level. The result would be displayed on the webpage. This helps the concerned authority to carry on a cleanup drive.

Graph 1: Dustbin is fully empty. Therefore, graph has no blue bar and it is fully white.



Graph 2: Dustbin is partially empty. Therefore blue bar is seen, Blue bar show how much dustbin is full. In below graph it can be seen that dustbin is 8.9 m full.

B. Outcomes

The following are the outcomes were observed as a result of implementing the product:

- Waste Level detection inside the dustbin.
- Transmit the information wirelessly to local authorities.
- The data can be accessed anytime and from anywhere.
- The real-time data transmission and access.
- Avoids the overflows of Dustbins

IV. DISCUSSION OF THE RESULT.

This IoT based waste management is extremely useful for smart cities and even the agricultural areas as equally. Even in the rural areas lack of garbage management has been leading to various diseases that have been taking shape in different aspects. Hence, the monitoring will widely help them take the necessary precautions and build health into rural areas. IT has been observed seen that in cities there are different dustbins located in the different area's and dustbins get overflown many times and the concerned people do not get information about this. The system is designed to solve this issue and will provide complete details of the dustbin located in the different areas throughout the city. The concerned authority can access the

knowledge from anywhere and anytime to urge the small print. Accordingly, they can take the decision on this immediately. In addition to added, advantages of being able to monitor the garbage bin over various areas, a feature has been incorporated where the overflowing bin location, when triggered could send out the information to the concerned authority. This particular feature further enhances the capability that this particular product or project brings to the table

V. FUTURE SCOPE.

The main objective is to take care of the extent of cleanliness within the city and form an environment that is best for a living. By using this technique we will constantly check the extent of the rubbish within the dustbins which are placed in various parts of the town. If a specific dustbin has reached the utmost level then the workers are often informed and that they can immediately take certain actions to empty it as soon as possible. The workers can check the status of those bins anytime on their mobile phones. This will convince be a really useful system if used properly. The system is often used as a benchmark by the people that are willing to require one step further for increasing the cleanliness in their respected areas. An ultrasonic sensor is getting used during this system to see the extent of garbage within the dustbins but in future various other sorts of sensors are often used with the ultrasonic sensor to urge more precise output and to require this technique to a different level. Now, this technique is often utilized in certain areas but as soon because it proves its credibility it is often utilized in all the large areas. As this technique also reduces manual work certain changes are often wiped out the system to require it to a different level and make it more useful for the workers and other people who are using it. In the future, a team is often made which can be responsible for handling and maintaining this technique and also to require care of its maintenances.

VI. CONCLUSION.

This project work is the implementation of smart garbage management system using the ultrasonic sensor, Raspberry Pi and Wi-Fi module. This system assures the cleaning of dustbins soon when the rubbish level reaches its maximum. If the dustbin isn't cleaned in specific time, then the record is shipped to the upper authority who can take appropriate action against the concerned contractor. This system also helps to watch the fake reports and hence can reduce the corruption within the overall management system. This reduces the entire number of trips of garbage pickup vehicle and hence reduces the general expenditure related to the rubbish collection. It ultimately helps to keep cleanliness in society. Therefore, the smart garbage management system makes the rubbish collection more efficient.

References.

A. Journal References

[1] IOT based Smart Waste Management (2017). Prof. Indu Anoop, Ayush Jaina, Shweta Pathak, Gauri Yadav ISSN (Online) 2278-1021 ISSN (Print) 2319 5940

[2] Smart Garbage Monitoring System ForWaste Management Prof.Norfadzila Mohd Yusof and Aiman Zakwan Jidin. ISSN 20179701098

[3] An IOT based Dynamic Garbage Level Monitoring System using Raspberrypi Prof Harshita A. Gawad, Suraj Kadam, Dona. ISSN - 2248-9622

[4] The Data from the dustbin can be connected to the cloud. Prof Harshita Chugh, Dushyant singh, Shahensha shaik, Ashwani Singla e-ISSN: 2395-0056 p-ISSN: 2395-0072

Convo Hand – Smart Glove

Aditi Naik,Vanshika Nair, Nidhi Mishra, Abhinav Dubey
Thakur College of abEngineering and Technology, Mumbai, India
aditinaik0604@gmail.com, vanshikanair46@gmail.com, missmishra1999@gmail.com, abhinavdube98@gmail.com

Abstract-**In this paper we represent smart glove for deaf and dumb patient. About nine billion people in the world are deaf and dumb. The communication between a deaf and normal visual people. This creates a very little room for them with communication being a fundamental aspect of human life. The blind people can talk freely by means of normal language whereas the deaf-dumb have their own manual-visual language known as sign language. The smart wearable glove will solve many issues related to different fields like healthcare, defense sector, automation, public safety, control systems and many more. Once, it comes in public domain then it will truly change the life of many and will encourage such innovative projects and ideas so that more and more similar product is introduced for the benefit of the society.**

Keywords- Smart Glove; Solutions for Deaf and normal visual people; Gesture Vocalizer App; ConvoHand;

I. INTRODUCTION

The internet of things (IOT) is a computing concept that describes the idea of everyday physical objects being connected to the internet and being able to identify themselves to other devices. The IOT is significant because an object that can represent itself digitally becomes something greater than the object by itself.

General, deaf people have difficulty in communicating with normal people. Even those who do speak aloud typically have a "deaf voice" of which they are self-conscious and that can make them reticent. Hence there is a need of the system which recognizes the different signs and conveys the information to the normal people.

The ConvoHand glove is a normal glove with buttons along the length of each finger and the thumb. Buttons are used instead of flex sensors to make the product cheaper. The button output a stream of data that varies with ASCII code. The output from the button are analog values, converted to digital and processed by using microcontroller and then it will be transmitted through Bluetooth wireless
communication, then it will be received in the receiver section and processed giving the output. The output is been shown in an app called CONVOHAND which is made through MIT App inventor-2, the responses are in text and voice. In this project buttons plays the major role.

Analysis is the third phase in software development lifecycle which aims in finding the requirements for the project which we are going to implement. Analysis means to understand the business needs and processes so that the project we implement becomes efficient. In system analysis we consider what the
system will need to take into consideration all the
points so that the stakeholders will have the right model built That signifies to collect the requirements which reflects the stakeholder's expectations. So, in this phase we spent considerable amount of time talking with the customers and understanding their requirements. In short, we can say that analysis phase means conquering requirements and analyzing them. So, in analyzing requirements we create the deadline and consider various ways to achieve it and hence the requirements are considered formal requirements.

II. OBJECTIVE

The main objective of this project is to help deaf and dumb people by removing communication barrier so that they are not restricted in a small social circle and are able to convey their feelings and emotions . This project also aims to solve the daily challenges faced by the people, who are unable to speak (dumb) or one who has recently undergone an accident and is unable to speak. It can also be used by elderly people, who find difficulty in speaking. Whenever they want. Also it would be helpful in education and health issues related to deaf and dumb people. Assertive technology products for special people to promote building an inclusive society. This work includes a voice based and text based interaction approach.

III. SCOPE

To achieve the objectives, the scope of this project is determined. With this wearable smart glove, dumb people or patient can easily communicate by just tapping the points on the glove by their thumb that results in 4 different commands that are both audible audio and image on any Android smartphone via an app. This can also be used in automation of day to day things like home appliances and many more. This can also be integrated in the field of automation where one can automate things without getting in direct contact of the machines/appliances that are to be automated or controlled.

IV. PROBLEM DEFINATION

Problems faced by the disable person regarding employment can be overcome by our method. So in the implemented work an intelligent microcontroller based system using Flex sensors is developed which is able to convert gesture into voice and text. Help a person to control his home appliances if he could not walk to

switchboard .In today's technology wireless gloves are not yet reliable because to be used as wireless, the gloves should have inbuilt battery and some electronics controller board which makes gloves heavier and may cause irritation. Thus wired equipment's are preferred for patients and partial disable people.

V. OBSERVATION MATRIX A

Observation

In the society, we observe that many problems
occurred while communicating with deaf and dumb people.

- Unable to convey the message
- Takes time to understand
- Time wastage
- Less security
- Less safety

B. Scouted challenges

In current technology we face the many problem.

- More reliable
- Automatic system
- Fast working
- Less time consumes
- Security
- Base planning

C. Top Five problems

There are many problems with these people. Problem like

- people like this can't get the employment.
- At least one person is required to understand the person.
- Delay in understanding may lead to unexpected problems.
- Patient in hospital always requires an assistant.
- People may take the advantage of their disability.

VI. FEATURES OF THE PROJECT

It can solve the daily difficulties suffered by the people, who are cannot speak (dumb) or one who has recently undergone an accident and is not in a condition to speak about this health to the doctor or express his feeling to his family. It can also be used by elderly people, who find difficulty in talking due to their bad heath conditions.

This can also be integrated in the field of automation where one can automate things without getting in direct contact of the machines/appliances that are to be automated or controlled.

It has wide application in defense sector and war equipment, where solider will not put their life in danger and can control the artillery or explosive without coming in direct contact of it from a safe distant zone. This smart glove can also work as a controller for Wheel chair and stuffs that are wheeled and needs to be controlled by gestures in some critical or required situation. So, this glove can easily communicate by just tapping the points on the glove by their thumb that results in twelve different expressions/commands in text, image and audio format on any android smartphone via an app. But, its application possibilities are more in different fields.

VII. SYSTEM REQUIREMENT

A. Software

There are two main software used in developing this
project which are Arduino software and MIT App Inventor. Arduino software used to write the program using C language and Arduino language. It is used to compile and install the program into the microcontroller. It will display the value of button and make the project easier to be tested.

The MIT App Inventor, online app development application software. The application has two parts.

1. Screen Layout.

2. Blocks.

The MIT App Inventor development environment is supported for Mac OS X, GNU/Linux, and Windows operating systems, and several popular Android phone models.

B. Hardware

There are five main hardware in this projects.

1. Arduino UNO
2. Bluetooth module HC-05
3. Jumper wires
4. Glove and Metallic Buttons
5. Soldering Iron

VIII. IMPLEMENTATION FOR METHODOLOGY

A. Flow of the project implementation:

1. Connect the battery to the Arduino Uno.
2. on the Arduino and Bluetooth.
3. Open ConvoHand app and touch any one button make the app active.
4. Let the required person wear the glove and use the glove as it is required.

B. Block Diagram:

The figure 4.2 shows all the connections from the glove and the Bluetooth module HC-05 to the Arduino Uno. The color of the wire shown the figure 4.2 is identical to the wire in the working model.

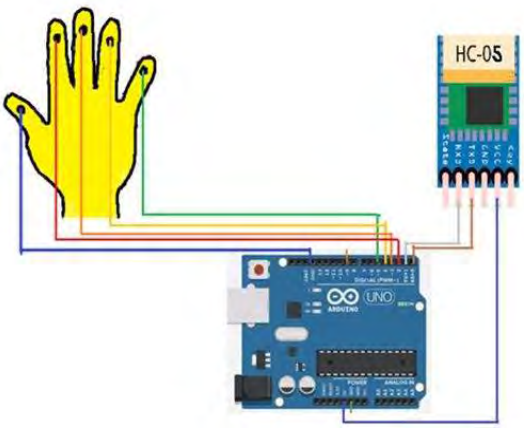


Fig1. Block Diagram of working model

C. Arduino UNO:

Arduino Uno is the main part of this project which it
act as microcontroller. Since the price is very cheap and easy to use, it is very good for the development of new products or interesting projects. Coming with
14 digital pin, 6 analog pin, 16 MHz crystal oscillator and USB connection, it is suitable to be used as a
microcontroller. Based on ATmega328, this microcontroller board can simply connect to computer using USB cable or using adapter to connect with battery to get started.

Fig 2. Arduino UNO

D. Bluetooth Module HC-05:

The Bluetooth module HC-05 is a MASTER/SLAVE
module. By default the factory setting is SLAVE. The Role of the module (Master or Slave) can be configured only by AT COMMANDS. The slave modules cannot initiate a connection to another Bluetooth device, but can accept connections. Master module can initiate a connection to other devices.
The user can use it simply for a serial port
replacement to establish connection between MCU
and GPS, PC to your embedded project, etc.

Fig3. Bluetooth Module HC-05

E. Glove and Metallic Button:

This project has been made by using plastic and rubber glove. We can also use woolen glove as per
convince. Metallic buttons has been attached to this glove in order to make the connection. On the backside of buttons we have solder electrical wires so that message can be passed using electrical medium.

Fig4. Glove and the Metallic buttons

F. Android App -MIT App Inventor:

The MIT App Inventor, an online app development
application where developer cans drag- drops the layout of Android app as the requirement of the developer. The MIT App Inventor project seeks to democratize software development by empowering all people, especially young people, to move from technology consumption to technology creation. The app can be cloned to an Android phone using MIT AI2 Companion that can be downloaded from Android Play Store, showing how exactly it looks on an Android phone. If you don't have an Android phone, you can build your apps using the Android emulator, software that runs on your computer and behaves just like the phone. The MIT App Inventor development environment is supported for Mac OS X, GNU/Linux, and Windows operating systems, and several popular Android phone models. Using MIT App Invertor, developer can make his own Android app using Build option using a QR code using .apk.

Fig5. Android App

G. CONVOHAND APP:

Fig6. Display page of the CONVOHAND App

This is the first page of app. The first line of page include title of the app the second line shows virtual voice which means it will convert the message in voice message for users. A different command has different voice messages.

Fig7. Command 1

This page shows the result of command touched by the user. 8 128 128 128 is the code value with
convert the number from Arduino to app and helps to display the correct message.

Fig8. Command 2

IX. RESULT ANALYSIS

As result of the project we are able to achieve wearable smart glove. Which is not only useful for deaf and dumb but also for people who don't get any employment. This smart glove is connected with android app which can be get transmitted through and android phone. Cost of this glove is very less as compare to other replacement available in market. Rather than using any sensor this glove has been developed with the help of metallic buttons which is very cheap.
Thus, the smart wearable glove will solve many issues related to different fields.
Once, it comes in public domain then it will truly change the life of many and will encourage such
innovative projects and ideas so that more and more similar product are introduced for the benefit of the
society.

Fig9. Result of working model

X. OUTCOMES PHASE 1:

i. Analysis:
Creating the synopsis, literature survey and feasibility
study for proposed solution which will act as an input to the design phase.

ii. Planning:
Proposing Statement of work, scope definition and
scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done.

iii. Designing:
Proposing the design architecture for the prototype.

iv. Implementation:
Implementation of Arduino UNO with the help of
code running in the Arduino
IDE environment.

PHASE 2:

i. Testing:
Testing was done by sending messages multiple
times to the GSM module.The testing of the model successes are mentioned in the above figures.

ii. Deployment:
When a real-time notice is sent to the GSM Module,
it is extracted and sent to Arduino, which is displayed on the LCD Screen. Thus, the model works efficiently.

XI. CONCLUSION

With CONVOHAND: Smart Glove we are able to help disable person in their daily work. As it is difficult to monitor and understand disable person, with ConvoHand he/she can easily communicate with others to express their feelings. This is achieved by making the glove consisting a button on each finger carrying a specific message with could be pressed and displayed on the screen to the concerned person. These messages are going to be displayed via app named "CONVOHAND" which was build on the application known MIT APP INVENTOR. This app has been connected with Arduino by Bluetooth module. This project is really helpful in medical field as well as in the customer care field.

References

[1]. Pallavi verma, S.L.Shimi and S. Chatterji, "Design of Smart Gloves", International Journal of Engineering Research & Technology (IJERT),Vol.3, Issue 11, November-2014, pp.210-21.

[2]. Khushboo Kashyap, Amit Saxena, Harmeet Kaur, Abhishek Tandon, Keshav Mehrotra, "Digital Text And Speech Synthesizer Using Smart Glove For Deaf And Dumb", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) ISSN: 2278 – 909X Volume
6, Issue 5, May 2017.

[3]. Arsh.A.Ghate1, Vishal.G.Chavan2, "Smart Gloves for Blind", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-
0056 Volume: 04 Issue: 12 | Dec-2017 www.irjet.net p-ISSN: 2395-0072.

[4]. Divyashree Merthia, Ayush Dadhich, Bhaskae Varma and Dipesh Patidar Porrnima Institute Of Engineering And Technology, Jaipur Rajasthan India "A Speaking Module For Deaf And Dumb". SSRG International Journal of Electronics and Communication Engineering (SSRG-IJECE) - Volume 3 Issue 8

Self-stabilizing spoon for Parkinson's sufferers with pulse sensor

Mukesh Choudhary, Chinmay Dali, Jignesh Dhakad, Sunnykumar Bhagat
TCET, MUMBAI
mukeshchoudhary464@gmail.com, chinmaydali395@gmail.com, jignesh.dhakad@gmail.com, bhagatsunny1998@gmail.com

Abstract— **Parkinson's disease (PD) is a progressive disorder that affects nerve cells in the brain responsible for body movement which accompanied by aberrant behavior in the patients .Other movement disorders include cerebral palsy, ataxia, and Tourette syndrome. Nerve cells, or neurons, are responsible for sending and receiving nerve impulses or messages between the body and the brain. It happens when the patient tries to move or stay still. The symptoms of Parkinson's disease develop gradually. They often start with a slight tremor in one hand and a feeling of stiffness in the body. The tremors first occur in the patient's hands which causes and of all the problems that come along the worst is when they try to eat . The aim of this project is to develop a supporting unit for the patients suffering from Parkinson's disease to help them create a system for stabilizing and leveling handles, designed to help people with hand tremor or limited hand and arm mobility thereby retaining dignity, aplomb, and self-subsistence during their eating process. A prototype of a "self-stabilizing spoon" was constructed to work under real circumstances and intended to be a complement for people who are in need of assistance during their eating process. To make this possible, a sensor with gyroscopes combined with accelerometers was used to identify which direction the device's handle was being tilted, as well as how fast its position was changed. Two servo motors were placed orthogonally to each other to establish a system that operates in 2D[two dimensions]. With this setup, the spoon was intended to maintain its spoon bowl in a horizontal position as they were intended to be.Thereby patient can easily eat their food. Experimental results of the spoon showed interesting results, performance with some unconventional limitations.**

I. INTRODUCTION

Parkinson's Disease is one of the most common neurological disorder and is defined as "a progressive disease of the nervous system marked by tremor (rapid and unintentional movements), muscular rigidity, and slow, imprecise movement, mostly impact middle-aged and elderly people. the world population, over 60 years old, lives with PD . It is a intensifying disorder that is caused by degeneration of nerve cells in the, which controls the movement.Over time, other symptoms develop, and some people will have dementia. Beginning periods of the ailment, the side effects are shaking, unbending nature,trouble with strolling and gradualness of development. The Demeanor and debating issues may likewise happen for this PD.The Parkinson's disease is analyzed by specialist by making inquiries about the indications and they will tests that show how well the patinas nerves are functioning. There are no labs or blood tests for the analysis, at times, specialist may have attempt a medication.The aim of this project is to produce a

stabilizing spoon that will compensate for unintended motions, such as tremors. With a low budget, the goal is to make a highly efficient prototype that consists mainly of a microcontroller and servo motors To make this possible, a sensor with gyroscopes combined with accelerometers was used to identify which direction the device's handle was being tilted, as well as how fast its position was changed. Two servo motors were placed orthogonally to each other to establish a system that operates in 2D[two dimensions].There are several different technological assisting spoons on the market today, but the products are unfortunately quite expensive [liftware.com, 2016]. With results from this project, further research could be able to continue developing cheaper and better stabilizing spoons.

II. PROBLEM DEFINITION

Design a self-adjusting spoon for patients having parkinson;s disease, so that during tremor they are able to eat food without spilling the food.Along with spoon there is a pulse rate sensor that records the pulse rate everytime the spoon is held.

III. BACKGROUND

Pharmacological therapy is an option to mitigate tremors but it has its own aftereffects. There are other prying technologies e existing like "Deep Brain Stimulation" is a process which is neurosurgical which utilizes microelectrodes to electronically stimulate definite areas of the brain. There are other existing technologies like gyroglove directly stabilise the hands of someone suffering from tremors. There are companies such as 'GYENNO' and 'Liftware' who are manufacturing "Smart Spoon" or "self-stabilizing spoon" which stabilizes 85% of tremors, although the majority of users are very happy, some find it difficult to use since their handles are very bulky and putting pressure might damage the product.

IV. IMPORTANCE

The requirement of the self-stabilizing spoon is because of the increase in the number of Parkinson's disease sufferers and as a result they have problem while doing daily activities. With the help of self-stabilizing spoon, the

problem of tremor can be solved up to a great extent. Importance of Self-stabilizing spoon are as follows:-In this project, we aim at designing a handle for the Parkinson patients to eat on their own.The spoon will be light and comfortable enough to be used by the patient.Support system for disabled people so that they won't worry about eating.The spoon is only an attachment to a stabilizing handle, so you have a selection of an everyday spoon, soup spoon and a fork attachment.

V. OBJECTIVES AND SCOPE

The objective of this project is to investigate how an Arduino microcontroller can be implemented to help people with impaired motor skills. Constructing a self-stabilizing spoon for Parkinson's sufferers. Every time when the spoon is used the patients pulse rate is also checked and can be seen anywhere. The spoon has the ability to move and rotate around x-axis and y-axis along with the spoon the pulse rate can be checked and displayed on thingspeak api. Using this spoon, it is helpful to a great extent to patients as well as the doctors.

VI. PHASES

Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: Getting strong knowledge of the project title and doing exploration on it, we will get our description. we will first study about it and do some research on the project for our better understanding and also getting a rough depiction about what would be our problem definition for this particular project.

Designing: Then we will build the design of the project and agreeing to that, will list down all the necessities needed for the creation for the prototype of our project. Implementation: After gaining the requirements, we first developed the connections of the hardware components (Arduino, MPU6050, Servo motor) and then added the coding part.

Phase 2: Testing and Deployment.

Testing: After the model is ready we will first associate the hardware with the allotted code and then we will check if it supports the machine or not. If not we will resolve the issues relating to it and will crisscross again.

Deployment: After completion of integration and testing of project, real time testing and operation of the system will be done.

VII. METHODOLOGY

Keeping all the data in the mind, regarding the architecture of the device the placement of our Arduino Uno and the connection of two consecutive servo motors along with each other is such that the horizontal shaft of one is connected with the other shaft causing the spoon to rotate. Both of them are connected to digital pins of Arduino Nano. MPU6050 sensor detect the tremor and roll angle of the vibration of the patient's hand. Simultaneously the pulse rate sensor will also count the pulse rate and display it on the Thingspeak api.

VIII. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware

- 1 Arduino Uno
- 1 MPU6050 gyroscopic sensor
- 2 servo motor
- Nodemcu
- 1 pulse rate sensor

Software

- Arduino IDE
- Thingspeak api

IX. CIRCUIT DIAGRAM

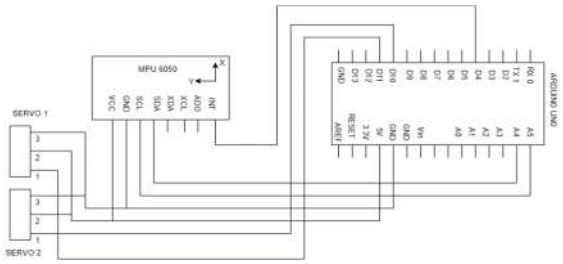


Fig. Circuit Diagram

This circuit diagram represents the connections of Arduino with Mpu6050 and servo motor. The pulse sensor is connected to nodeMCU so that it can send readings to thingspeak api over wifi.

X. OUTPUT

When we move or shake our hand ,the spoon rotates to opposite direction and stabilizes itself, parallelly pulse is checked

XI. RESULT

When activated the spoon is set to 180 degrees to hand level.on hand movement the servo motors adjust in horizontal as well as vertical direction.

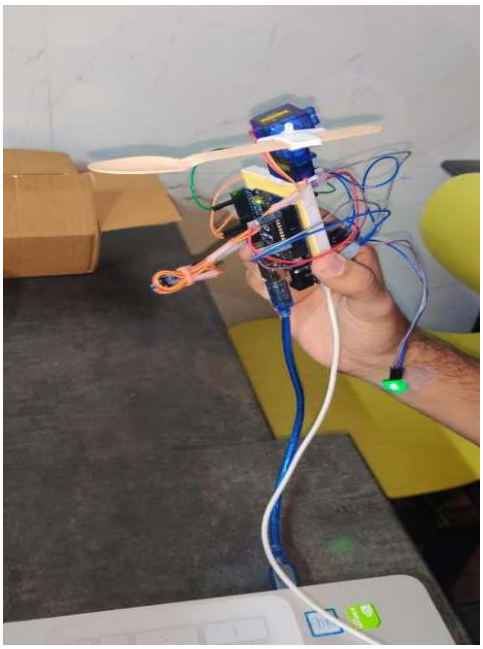


Fig. self-stabilizing spoon Kit

XII. OUTCOMES

Arduino Uno is connected to two servo motors for rotational movement, a MPU 6050 gyroscopic sensor to determine the pitch of the spoon and a spoon. After uploading the code on Arduino Uno, the servo motors adjust itself at a position of 90 degrees. When a movement of hand is encountered by the system, the gyroscopic sensor determines the direction and the angle of movement and sends the data to Arduino uno which in turn sends data to servo motor to react in the opposite direction to the reaction of hand. The spoon is adjusted in such a way that it always faces the upper direction irrespective of the movement. The pulse sensor which is attached to the hand and connected to the NodeMCU, senses the pulse rate and sends the data to the thingspeak api. The system is capable of Performing both tasks at same time.

XIII. RESEARCH GAPS

Every project has some gaps in it. No project is 100% efficient. There are some gaps:

Since our project uses two Arduino boards, one Arduino Uno and one Esp8266 Nodemcu, instead only one board can be used for both spoon and the pulse rate sensor. The code for Arduino can be optimized and high sensitivity and power servo motor can be used In order to React to the motion of hand.

XIV. CONCLUSION

We were able to stabilize the spoon in 3 dimensions. The main objective of the project is for Parkinson's disease which is one of the common neurodegenerative disorder. Due to this disorder the patient suffers from severe tremor related issue. When the person focuses on a particular body part he/she suffers severe tremor in that body part. Due to this the person is not able to hold things, eat properly and therefore suffers a lot of problem.To overcome this problem we designed an Self adjusting Parkinson's spoon for the Parkinson's sufferer while eating which help to control the tremor to a great extent. This device can be used as a rehabilitation or supportive device as it does not completely removes the tremors. Along with spoon the pulse sensor also checks the pulse rate and gives the results on Thingspeak api.

References

[1] https://www.hackster.io/leesanggyu/self-stabilizing-spoon-for-parkinson-s-sufferers-7bc5a7

[2] https://www.diva-portal.org/smash/get/diva2:1200521/FULLTEXT01.pdf

[3] https://www.ijraset.com/fileserve.php?FID=22713

Smart Irrigation System

Vaibhav Tiwari, Saurabh Yadav, Shreya Patil

Thakur College of Engineering and Technology

Mumbai, India

vt65760@gmail.com, sy083844@gmail.com ,shreyapatil1815@gmail.com

Abstract—**Proposed system for Smart Irrigation System, automation of farm irrigation and soil moisture control by Arduino using soil moisture sensor and L293D module. This automatic irrigation system senses what the moisture content of the soil is and it automatically switches the pump on when the power is on. A proper usage of irrigation system is very necessary because there is a shortage of land reserved water due to lack of rain, and the spontaneous use of water, and as a result large amounts of water goes waste. This is exactly the reason why we use this automatic plant watering and soil moisture level detecting and monitoring system. This system is very useful in all climatic conditions. India is an agriculture based country. A vast majority of people in India are completely depended on the agricultural harvesting. Agriculture is a source of employment of a majority of Indians and has great impact on the economy of the country. In dry areas or in areas lacking rainfall, irrigation becomes difficult. So, it needs to be automated for proper watering of crops, to be done by remote handling by the farmer. When the soil goes dry or if the moisture content in the soil is less than that which is advised, the pump will start watering the crops.**

Keywords: Smart Irrigation, Sensors, Bluetooth communication, Android.

I.INTRODUCTION

The main aim of this proposed system was to provide water to the plants or gardening automatically using microcontroller (Arduino Uno). One can automatically water the plants, say, when said person are going out on vacation and there is no possible way to water the plants by physical presence, and we don't want to bother the neighbors. Sometimes the neighbors water the plants excessively and the plants end up dying anyway. There exist devices which are timer based in India which water the soil on set intervals. They fail to sense the soil moisture and the immediate surrounding temperature to recognize if the soil actually needs watering or not. Assimilation is that, the artificial application of water to the land or soil. It is used to support in the developing of agricultural crops, preservation of landscapes, and re vegetation of disturbed soils in dry areas and during periods of scarce rainfall. When a sector comes on, the water flows through the adjacent lines and finally ends up at the irrigation electrode (drip) or mechanical device heads. Several sprinklers have pipe thread creeks on the lowest of them that allows a fitting and also the pipe to be connected to them. The sprinklers are usually attached to the top of the head flush with respect to the ground surface. As this dripping method reduced not only huge water losses but it also became a popular method by reducing the labor cost and increasing the profits and harvests. When the components are activated, all the components will read and give the output signal to the controller, and the information will be displayed to the user (farmer).

II.SCOPE OF THE PROJECT

The Smart Irrigation System has an extensive scope to automate and industrialize the complete irrigation system. The proposed system aims and revolves around building an IoT based Irrigation System using Arduino UNO Module and DHT11 Sensor. It will not only automatically irrigate the water based on the moisture level in the soil but also send the data to Blynk App to keep track of the land condition. The System will consist a water pump which will be used to sprinkle water on the land depending upon the land environmental condition such as Moisture, Temperature, Humidity.

III.PROPOSED SYSTEM

An automatic plant watering system using Arduino microcontroller UNO R3 is programmed such that it gives the interrupt signals to the motor via the motor driver module. The soil sensor is connected and linked to the A0 pin of the Arduino board which senses and makes a note of the moisture content existing in the soil. Whenever the soil moisture content values goes down, the sensor senses the humidity change, giving signal to the microcontroller so that the pump (motor) can be activated. This theory can be used successfully for automatic plant watering system. The circuit includes an Arduino UNO board, a soil moisture sensor, a 5V motor pump, a Motor driver L293D (IC1), motor driver IC to run the water pump. One can control the Arduino board using a 5V-9V plug in adaptor or the best, solar panel. A separate 5V-9v battery for the pump motor is a must.

3.1 FLOWCHART

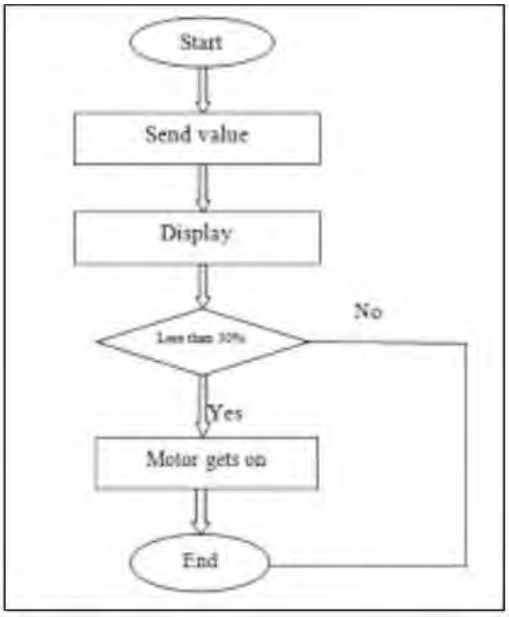


Fig: Flowchart for Smart Irrigation System

3.2 SYSTEM DESIGN

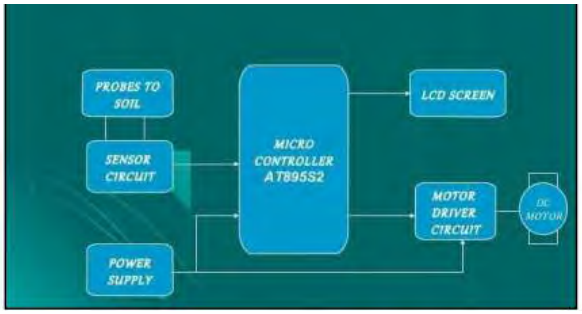


Fig: System Design Diagram for Smart Irrigation System

3.3 WORKING OF THE CIRCUIT

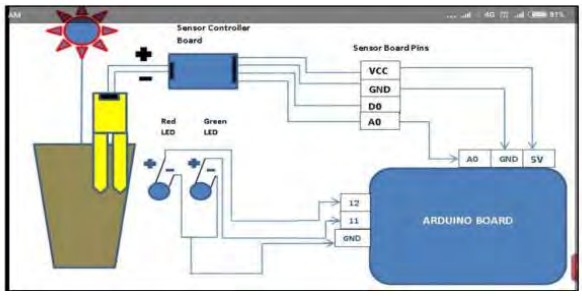


Fig: Working of the Circuit for Smart Irrigation System

IV.COMPONENTS AND DESCRIPTION

i.Arduino Uno –

The Arduino Uno R3 (also called ATmega328) is a dual-inlinepackage (DIP) microcontroller AVR microcontroller. Overall, it has 20 I/O digital pins. Arduino computer programs are easily loaded to full on the Arduino Uno R3.

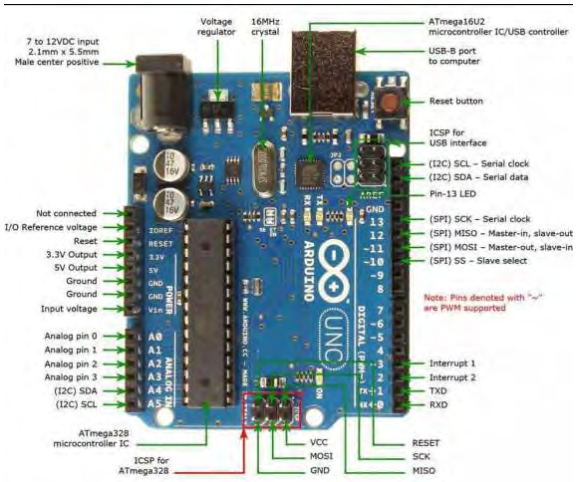


Fig: Arduino Uno

ii.Breadboard –

Initially it was literally a bread board, it was a polished piece of wood used for cutting and slicing bread. In the 1970s the solder less breadboard became available; nowadays the term "breadboard" is commonly used to refer to these.

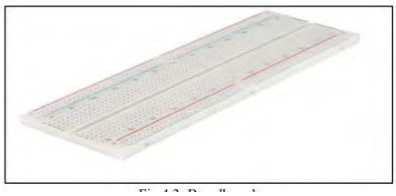


Fig: Breadboard

iii.Jumper Wires-

Jumper wires are merely wires that have connector pins at each end, which allows them to be used to connect two points to each other without the need of soldering. Jumper wires are naturally used with breadboards and other prototyping tools to make it easy to change a circuit as and when needed according to the requirement.

Fig: Jumper Wires

iv.Soil Moisture Sensor-

Soil moisture sensors guesses and gives a rough estimate of the soil water content based on the dielectric constant of the soil. The dielectric constant can be perceived as the soil's capability to transmit electricity. The dielectric constant value of soil rises as the water content level of the soil increases. This reaction is because the dielectric constant value of water is much superior than the other soil apparatuses, including air. Thus, measurement of the dielectric constant gives a probable approximation of

water content. It consists of a pair of electrodes to measure and note the resistance of the soil.

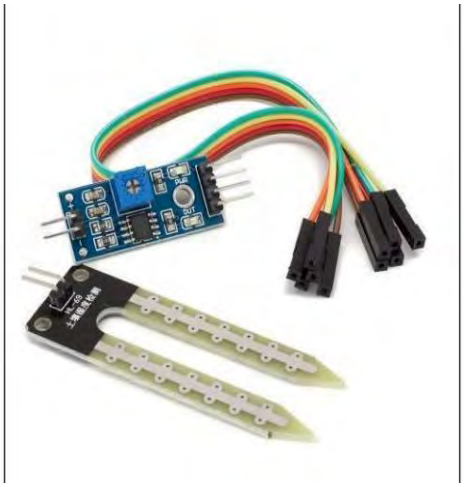


Fig: Soil Moisture Sensor

V.RESULTS AND DISCUSSIONS

a. Project Working:

Fig: Project working

The project working goes as: the soil moisture sensor, senses the moisture level that we have set on the Blynk App, and if the moisture content is less than the set moisture level, then the motor in the water pump will automatically start and start pumping out water through the pump, and it will continue to do so until the said moisture level is required. On the other hand, if the soil moisture sensor picks up a level which exceeds the set level, then the motor in the water pump will not start and there will be no pumping of water, thus not overwatering the plant, leading to its death.

b. Blynk App interfacing:

Fig: Blynk App Interfacing

On the Blynk App, a new project is created and the user can set the limits of the humidity, temperature and moisture. Through this app, it is also possible for the user to remotely water the plants with the mere click of a button and the user doesn't have to necessarily be physically present for watering the plant. Thus giving an ease and comfort.

VI.ANDROID MOBILE APPLICATION

Android is used to develop mobile application for automatic irrigation. Android's user interface is mainly based on direct manipulation using touch gestures that loosely correspond to real-world actions, such as swiping, tapping to manipulate on-screen objects, along with a virtual keyboard for text input. The data of the sensor and onset values are stored in the local memory of the mobile. The user can read and know the sensor data, and can regularize the system into an automatic mode so that the system can automatically switch on the motor depending on the sensor data value and previously set onset value. The user can also set the system to manual mode and he himself has the ability to decide the switching on and off of the motor. And he can get the notifications concerning the water level in tank so that he can take and make alternative decisions when there is shortage of water in tank.

VII.FUTURE SCOPE

The system can be interfaced with cloud, using Raspberry pi 3b or Bolt IoT. Further GSM module can be interfaced with it for SMS features. This worked quite perfectly but some parts of the secondary aim was left considering the fact of us being not expert in programming. The system built works only on one resistance (default) value measured by the sensor thus moistening the field always by the same amount for all the plants. Due to difficulties and complications arising in the programming, different resistance values cannot be set for diverse plants to provide altered level of moisture which can be the future development for the proposed system.

VIII. CONCLUSION

The main objective of this proposed smart irrigation system is to make it more innovative and ground-breaking, user friendly, comprehensible, tangible, time saving and more efficient than the existing systems available. Measuring four parameters such as soil moisture content and the system also includes intruder detecting system. Due to the updates given by the app, the farmer can recognize about crop field nature at any given time, anywhere. The application of agriculture networking technology is in need of the modern agricultural development, and this is the future of agricultural development. After building the agricultural water irrigation system hardware and analyzing and researching the network hierarchy features, functionality and the corresponding software architecture of precision agriculture water irrigation systems, actually applying the internet of things to the highly effective and safe agricultural production has a significant impact on ensuring the efficient use of water resources as well as ensuring the efficiency and stability of the agricultural production. With more advancement in the field of IoT expected in the coming years, these systems can be more efficient, much faster and less costly. In the Future, this system can be made as an intelligent system, where in the system predicts user actions, rainfall pattern, time of harvest, animal intruder in the field and collaborating the information through advanced and cutting-edge technology like IoMT can be implemented and employed, so that the whole agricultural system can be made self-regulating and independent of human operation and in turn quality and huge quantity yield of crops can be obtained.

References

[1] https://www.watelectronics.com/automatic-plant-irrigation-system-circuit-and-its-working/

[2] IoT Based Agriculture Monitoring and Smart Irrigation System Using Raspberry Pi Mrs.T.Vineela1, J. NagaHarini2, Ch.Kiranmai3, G.Harshitha4, B.AdiLakshmi5, Volume: 05 Issue: 01 | Jan2018.

[3] IOT based Smart Irrigation System Srishti Rawal Department of Computer Science, Volume 159 – No 8, February 2017.

[4] Smart Agriculture Based on Cloud Computing and IOT Fan TongKe Modern Education Technology Center of Xi'an International University, Shaanxi,710077, China, fantongke@126.com

[5] REVIEW PAPER BASED ON AUTOMATIC IRRIGATION SYSTEM BASED ON RF MODULE Ms. Deweshvree Rane PG Scholar - VLSI, Sevagram, Wardha, India

[6] Priadarshani K. AN IoT BASED SMART IRRIGATION SYSTEM, 2017

Text – To – Speech Synthesis (TTS)

Advait Singh, Aditya Singh
TCET MUMBAI
singhadvait16@gmail.com, iamsinghaditya@gmail.com

Abstract – Speech is one of the oldest and most natural means of information exchange between human. Over the years, Attempts have been made to develop vocally interactive computers to realise voice/speech synthesis. Obviously such an interface would yield great benefits. In this case a computer can synthesize text and give out a speech. Text-To-Speech Synthesis is a Technology that provides a means of converting written text from a descriptive form to a spoken language that is easily understandable by the end user (Basically in English Language). It runs on JAVA platform, and the methodology used was Object Oriented Analysis and Development Methodology; while Expert System was incorporated for the internal operations of the program. This design will be geared towards providing a one- way communication interface whereby the computer communicates with the user by reading out textual document for the purpose of quick assimilation and reading development.

Keywords: *Communication, Expert System, FreeTTS, JAVA, Speech, Text-To-Speech*

I. INTRODUCTION

Voice/speech synthesis is a field of computer science that deals with designing computer systems that synthesize written text. It is a technology that allows a computer to convert a written text into
speech via a microphone or telephone. As an emerging technology, not all developers are familiar with speech technology. While the basic functions of both speech synthesis and speech recognition takes only minutes to understand, there are subtle and powerful capabilities provided by computerized speech that developers will want to understand and utilize.
Automatic speech synthesis is one of the fastest developing fields in the framework of speech science and engineering. As the new generation of computing technology, it comes as the next major innovation in
man-machine interaction, after functionality of Speech recognition (TTS), supporting Interactive Voice Response (IVR) systems.
The basic idea of text-to-speech (TTS)
technology is to convert written input to spoken output by generating synthetic speech. There are several ways of performing speech synthesis:
1.Simple voice recording and playing on demand;

2.Splitting of speech into 30-50 phonemes (basic linguistic units) and their re-assembly in a fluent speech pattern; The maximization of these two criteria is the main development goal in the TTS field.3.The use of approximately 400 diaphones (splitting of phrases at the centre of the phonemes and not at the transition).
The most important qualities of modern speech synthesis systems are its naturalness and intelligibility. By naturalness we mean how closely the synthesized speech resembles real human speech. Intelligibility, on the other hand, describes the ease with which the speech is understood. The maximization of these two criteria is the main development goal in the TTS field.

II. OBJECTIVES OF THE STUDY

The general objective of the project is to develop a Text-to-speech synthesizer for the physically impaired and the vocally disturbed individuals using English language. The specific objectives are:
1. To enable the deaf and dumb to communicate and contribute to the growth of an organization through synthesized voice.
2. To enable the blind and elderly people enjoy a User-friendly computer interface.
3. To create modern technology appreciation and awareness by computer operators.
4. To implement an isolated whole word speech synthesizer that is capable of converting text and responding with speech
5. To validate the automatic speech synthesizer developed during the study.

III. SCOPE OF THE STUDY

The study is focused on an ideal combination of a human-like behaviour with computer application to build a one-way interactive medium between the computer and the user. This application was customized using only one (1) word sentence consisting of the numeric digit 0 to 9 that could be used in operating a voice operated telephone system.
Human speech is inherently a multi modal process that involves the analysis of the uttered acoustic signal and includes higher level knowledge sources such as grammar semantics and pragmatics. This project intends to focus only on the acoustic signal processing without the incorporation of a visual input.

IV. SIGNIFICANCE OF THE STUDY

This project has theoretical, practical, and methodological significance:
The speech synthesizer will be very useful to any researcher who may wish to venture into the "Impact of using Computer speech program for brain enhancement and assimilation process in human beings".

This text-to-speech synthesizing system will enable the semi-illiterates assess and read through electronic documents, thus bridging the digital divide. The technology will also find applicability in systems such as banking, telecommunications (Automatic system voice output), transport, Internet portals, accessing PC, emailing, administrative and public services, cultural centres and many others. The system will be very useful to computer manufacturers and software developers as they will have a speech synthesis engine in their applications.

V. TEXT – TO - SPEECH SYNTHESIS DEFINED

A speech synthesis system is by definition a system, which produces synthetic speech. It is implicitly clear, that this involves some sort of input. What is not clear is the type of this input. If the input is plain text, which does not contain additional phonetic and/or phonological information the system may be called a text-to-speech (TTS) system. A schematic of the text-to-speech process is shown in the figure 1 below. As shown, the synthesis starts from text input. Nowadays this may be plain text or marked-up text e.g. HTML or something similar like JSML (Java Synthesis Mark-up Language).

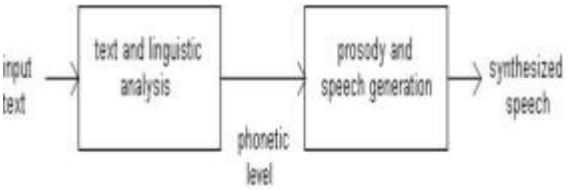


Figure 1: Schematic TTS

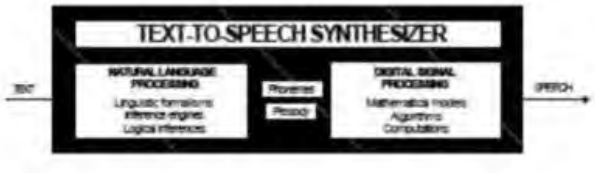


Figure 2: A general functional diagram of a TTS System.

5.1. Representation and Analysis of Speech Signals

Continuous speech is a set of complicated audio signals which makes producing them artificially difficult. Speech signals are usually considered as voiced or unvoiced, but in some cases they are something between these two. Voiced sounds consist of fundamental frequency (F0) and its harmonic components produced by vocal cords (vocal folds). The vocal tract modifies this excitation signal causing formant (pole) and sometimes anti-formant (zero) frequencies (Abedjieva et al., 1993). Each formant frequency has also amplitude and bandwidth and it may be sometimes difficult to define some of these parameters correctly. The fundamental frequency and formant frequencies are probably the most important concepts in speech synthesis and also in speech processing. With purely unvoiced sounds, there is no fundamental frequency in excitation signal and therefore no harmonic structure either and the excitation can be considered as white noise. The airflow is forced through a vocal tract constriction which can occur in several places between glottis and mouth. Some sounds are produced with complete stoppage of airflow followed by a sudden release, producing an impulsive turbulent
excitation often followed by a more protracted turbulent excitation (Allen et al., 1987). Unvoiced sounds are also usually more silent and less steady than voiced ones. Speech signals of the three vowels (/a/ /i/
/u/) are presented in time-frequency domain in Figure 3. The fundamental frequency is about 100 Hz in all cases and the formant frequencies F1, F2, and F3 with vowel /a/ are approximately 600 Hz,
1000 Hz, and 2500 Hz respectively. With vowel /i/ the first three formants are 200 Hz, 2300 Hz, and 3000 Hz, and with /u/ 300 Hz, 600 Hz, and 2300 Hz.

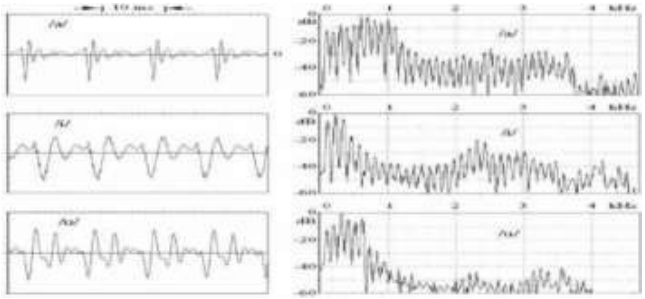


Figure 3: The time-frequency domain presentation of vowels /a/, /i/, and /u/.

VI. APPLICATIONS OF SPEECH SYNTHESIS

The application of synthetic speech is expanding fast whilst the quality of TTS systems is also increasing steadily. Speech synthesis systems are also becoming more affordable for common customers, making these systems suitable for everyday use. For example, better availability of TTS systems may increase employability for people with communication difficulties. Listed below are some of the applications of TTS system:

1. Applications for the Blind.
2. 2. Applications for the Deafened and Vocally Handicapped
3. Educational Applications.
4. 4. Applications for Telecommunications and Multimedia

5. Other Applications and Future Directions (e.g. Human-Machine Interaction)

VII. METHODOLOGY AND SYSTEM ANALYSIS

7.1 Analysis and Problems of Existing Systems

Existing systems algorithm is shown
below in Figure 4. It shows that the system does not have an avenue to annotate text
to the specification of the user rather it
speaks plaintext.
STOP

Figure 4: Algorithm of already existing systems

Due studies revealed the following inadequacies with already existing systems:
1. ***Structure analysis***: punctuation and formatting do not indicate where paragraphs and other structures start and end. For example, the final period in "P.D.P." might be misinterpreted as the end of a sentence.
2. ***Text pre-processing***: The system only produces the text that is fed into it without any pre-processing operation occurring.
3. ***Text-to-phoneme conversion***: existing
synthesizer system can pronounce tens of thousands or even hundreds of thousands of words correctly if the word(s) is/are not found in the data dictionary.

7.2 Expectation of the New System

It is expected that the new system will reduce and improve on the problems encountered in the old system. The system is expected to among other things do the following;
1. The new system has a reasoning process. 2. The new system can do text structuring and annotation.
3. The new system's speech rate can be
adjusted. 4. The Pitch of the voice can be adjusted.
5. You can select between different voices and can even combine or juxtapose them if you want to create a dialogue between them
6. It has a user friendly interface so that people
with less computer knowledge can easily use it
7. It must be compatible with all the vocal
engines 8. It complies with SSML
specification.

VIII. CHOICE OF METHODOLOGY FOR THE NEW SYSTEM

Two methodologies were chosen for the new system: The first methodology is Object Oriented Analysis and Development Methodology (OOADM). OOADM was selected because the system has to be represented to the user in a manner that is user-friendly and understandable by the user. Also since the project is to emulate human behaviour, Expert system had to be used for mapping of Knowledge into a Knowledge base with a reasoning procedure. Expert system was used in the internal operations of the program, following the algorithm of Rule-Based computation. The technique is derived from general principles described by researchers in knowledge engineering techniques (Murray et al., 1991; 1996).
The system is based on processes modelled in cognitive phonetics (Hallahan, 1996; Fagyal, 2001) which accesses several knowledge bases (e.g. Linguistic and phonetic knowledge bases, Knowledge bases about non-linguistic features, a predictive model of perceptual processes, and knowledge base about the environment).

9. Speech Synthesis Module

The TTS system converts an arbitrary ASCII text to speech. The first step involves extracting the phonetic components of the message, and we obtain a string of symbols representing sound- units (phonemes or allophones), boundaries between words, phrases and sentences along with a set of prosody markers (indicating the speed, the intonation etc.). The second step consists of finding the match between the sequence of symbols and appropriate items stored in the phonetic inventory and binding them together to form the acoustic signal for the voice output device.

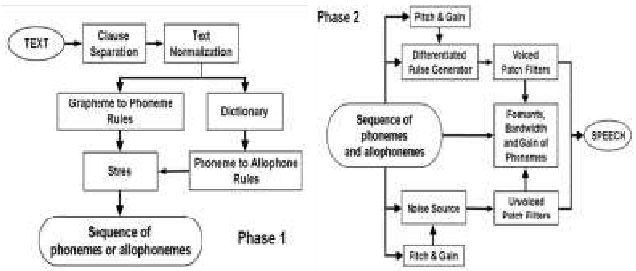
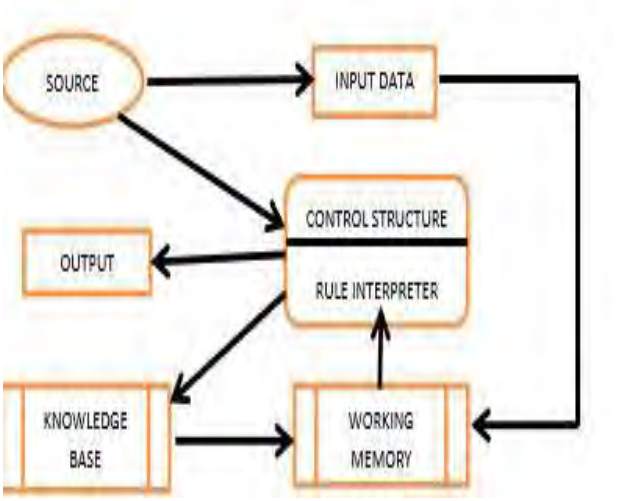


Figure 5: Phases of TTS synthesis process

To compute the output, the system consults
1. A database containing the parameter values for the sounds within the word,
2. A knowledge base enumerating the options for synthesizing the sounds.

Incorporating Expert system in the internal programs will enable the new TTS system exhibit these features:
1. The system performs at a level generally recognized as equivalent to that of a human expert
2. The system is highly domain specific.
3. The system can explain its reasoning process
4. If the information with which it is working is probabilistic or
fuzzy, the system can correctly propagate uncertainties and provide a range of alternative solution with associated



likelihood.

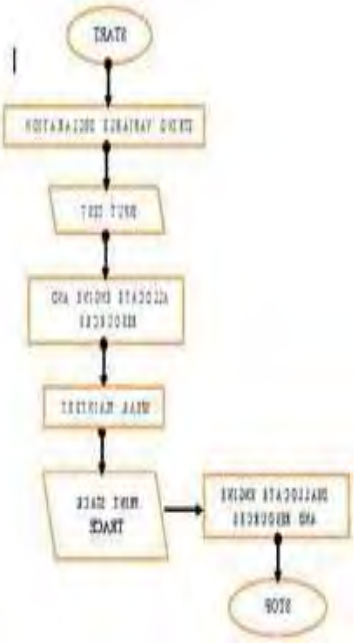


Figure 6: Data flow diagram of the Speech synthesis system Using Gane and Sarson Symbol

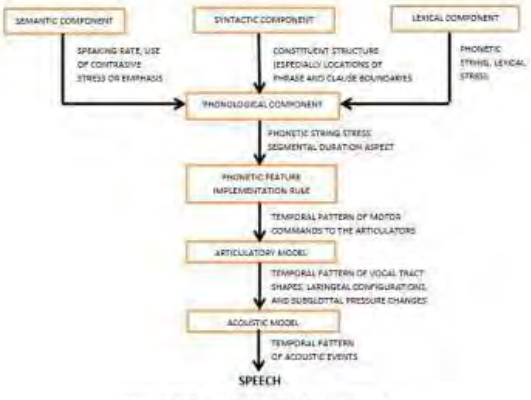


Figure 7: High Level Model of the Proposed System

User Interface (Source): This can be Graphical User Interface (GUI), or the Command Line Interface (CLI).

Knowledge Base (Rule set): FreeTTS module/system/engine. This source of the knowledge includes domain specific facts
and heuristics useful for solving problems in the domain. FreeTTS is an open source speech synthesis system written entirely in the Java programming language. It is
based upon Flite. FreeTTS is an implementation of Sun's Java Speech API.
FreeTTS supports end-of-speech markers. **Control Structures**: This rule interpreter inference engine applies to the knowledge base information for solving the problem. **Short term memory**: The working memory registers the current problem status and history of solution to date.

8.1 Choice of Speech Engine and Programming Language

The speech engine used in this new system was the FreeTTS speech engine. FreeTTS was used because it is programmed using JAVA (the backbone programming language of this designed TTS system). It also supports SAPI (Speech Application Programming Interface) which is in synchronism with the JSAPI (Java Speech Application Programming Interface). JSAPI was also the standardized interface used in the new system.
FreeTTS includes an engine for the vocal synthesis that supports a certain number of voices (male and female) at different frequencies. It is recommended to use JSAPI to interface with FreeTTS because JSAPI interface provides the best methods of controlling and using FreeTTS. FreeTTS engine enable full control about the speech signal. This new designed system provides the possibility to choose a voice between three types of voices: an 8 kHz, diphone male English voice named *kevin*, a 16 kHz diphone male English voice named *kevin16* and a16khz limited domain, male US English voice named *alan*. The user could also set the properties of a chosen voice: the speaking rate, the volume and the pitch.

A determining factor in the choice of programming language is the special connotation (JSML) given to the program. This is a java specification mark-up language used to annotate spoken output
to the preferred construct of the user. In
addition to this, there is the need for a language that supports third party development of program libraries for use
in a particular situation that is not amongst the specification of the original platform. Considering these factors, the best choice of programming language was **JAVA.** Other factors that made JAVA suitable were its dual nature (i.e. implementing 2 methodologies with one language), its ability to Implements proper data hiding technique (Encapsulation), its supports for inner abstract class or object development, and its ability to provide the capability of polymorphism; which is a key property of the program in question.

IX. DESIGN OF THE NEW SYSTEM

Some of the technologies involved in the design of this system includes the following:

Speech Application Programming Interface (SAPI): SAPI is an interface between applications and speech technology engines, both text-to-speech and speech recognition (Amundsen 1996). The interface allows multiple applications to share the available speech resources on a computer without having to program the speech engine itself. SAPI consists of three interfaces; The *voice text* interface which provides methods to start, pause, resume, fast forward, rewind, and stop the TTS engine during speech. The *attribute interface* allows access to control the basic behaviour of the TTS engine. Finally, the *dialog interface* can be used to set and retrieve information regarding the TTS engine. **Java Speech API (JSAPI):** The Java Speech API defines a standard, easy-to-use, cross-platform software interface to state-of-the-art speech technology. Two core speech technologies supported through the Java Speech API are speech recognition and speech synthesis.

Speech recognition provides computers with the ability to listen to spoken language and to determine what has been said. Speech synthesis provides the reverse process of producing synthetic speech from text generated by an application, an applet or a user. It is often referred to as text-to-speech technology.

9.1 Design of Individual Objects of the Program

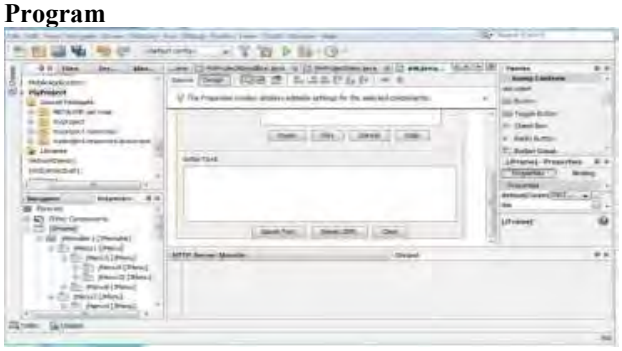


Figure 9: Netbeans Interface and program object manipulation

Figure 11: Overall view of the new system

9.2 Functions of the Abstract Classes

1. Menu Bar: This will have the function of selecting through many variables and File chooser system.
2. Monitor: Monitors the reasoning process by specifying the allocation process and de-allocation state
3. Voice System: This shows the different voice option provided by the system
4. Playable session: This maintains the timing of the speech being given out as output, and produces a speech in synchronism with the rate specified.
5. Playable type: This specifies the type of text to be spoken, whether it is a text file or an annotated JSML file
6. Text-to-Speech activator: This plays the given text and produces an output
7. Player Model: This is a model of all the functioning parts and knowledge base representation in the program
8. Player Panel: This shows the panel and content pane of the basic objects in the program, and specifies where each object is placed in the system
9. Synthesizer Loader: This loads the Synthesizer engine, allocating and de- allocating resources appropriately

X. CONCLUSION AND RECOMMENDATION

Synthesizing text is a high technology advancement and artificial formation of speech given a text to be spoken. With Text-to- Speech synthesis, we can actually mediate and fill in the lacuna provided by not fully exploiting the capabilities of some handicapped individuals. It's never been so easy to use a text-to-speech program, as just one click and your computer will speak any text aloud in a clear, natural sounding voice. Therefore, there is need to use Information Technology to solve the problem for the Before the use of the new system, proper training should be given to the users. This training can come in handy with proper tutor on how to handle JSML language and how to use it to annotate text for the proper output and emphasis.

9.2 Functions of the Abstract Classes

11. Menu Bar: This will have the function of selecting through many variables and File chooser system.
12. Monitor: Monitors the reasoning process by specifying the allocation process and de-allocation state
13. Voice System: This shows the different voice option provided by the system
14. Playable session: This maintains the timing of the speech being given out as output, and produces a speech in synchronism with the rate specified.

15. Playable type: This specifies the
type of text to be spoken, whether it is a text file or an annotated JSML file
16. Text-to-Speech activator: This plays the given text and produces an
output
17. Player Model: This is a model of all the functioning parts and knowledge
base representation in the program
18. Player Panel: This shows the panel and content pane of the basic objects in the program, and specifies where each object is placed in the system
19. Synthesizer Loader: This loads the Synthesizer engine, allocating and de- allocating resources appropriately

References

[1] Abedjieva et al. (1993): Acoustics and audio signal processing. http://www.ufh.netd.ac.za/bitstream/10 353/495/1/Mhlanathesis.pdf date: 21/07/12

[2] Allen, J., Hunnicutt, M.S., and Klatt, D. (1987). From text to speech – the MITalk system. MIT pre ss, Cambridge, Massachusetts.

[3] Hallahan (1996): Phonetics and Theory of Speech Production.http://www.indiana.edu/~acoustic/s702 /readfull.html date: 22/07/12

[4] I.R. Murray, J.L. Arnott, N. Alm and A.F. Newell (1991). A communication system for the disabled with emotional synthetic speech produced by rule. Proc. European Conference on Speech Communication and Technology 91.

[5] Abedjieva et al. (1993): Acoustics and audio signal processing. http://www.ufh.netd.ac.za/bitstream/10 353/495/1/Mhlanathesis.pdf date: 21/07/12

[6] Allen, J., Hunnicutt, M.S., and Klatt, D. (1987). From text to speech – the MITalk system. MIT pre ss, Cambridge, Massachusetts.

[7] Hallahan (1996): Phonetics and Theory of Speech Production. http://www.indiana.edu/~acoustic/s702 /readfull.html date: 22/07/12

[8] I.R. Murray, J.L. Arnott, N. Alm and A.F. Newell (1991). A communication system for the disabled with emotional synthetic speech produced by rule. Proc. European Conference on Speech

Arduino Based Real Time Air Quality Monitoring System

Priyanka Sharma, Riddhi Thanki, Saloni Singh
TE. Information Technology Thakur College of Engineering and Technology Mumbai, India
sharmapriyanka3585@gmail.com, iamriddhithanki@gmail.com, thakursaloni1999@gmail.com

Abstract— **Universal object interaction conceptualizes the idea of remotely connecting and monitoring real world objects (things) through the Internet. The air quality has become major issue from last decade due to the increase of pollutants in the atmosphere. Urbanization has boosted the number of air pollutants in numerous territories. This is damaging human health to a great extent. Thus, we developed an Arduino based system to measure the quality of air in the environment. The system uses MQ135 sensor to displays the real time data of the quality of air on LCD and Blynk mobile application using internet.**

Keywords— Air quality, MQ135 Sensor, Arduino, PPM.

I. INTRODUCTION

Air plays an important role in the ecosystem. Due to increase in the count of industries, vehicle, etc. various harmful gases are emitted in the atmosphere at large. The wellspring of air toxins can be regular or anthropogenic. Gases like Nitrogen Dioxide, Sulphur Dioxide, Carbon Dioxide, Carbon Monoxide and Ozone are the main cause of air pollution. These gases have harsh impact on human health. Air pollution causes various respiratory disorder, minor allergies, cerebral pain, aggravated asthma, eye and nose irritation and many more diseases. As indicated by WHO consistently 9 out of 10 individuals inhale air containing significant level contaminations.[1] Long time presentation to surrounding air results into passing. In view of study 4.2 million individuals bite the dust each year because of introduction to encompassing air contamination. Polluted air not only harms human but crops and animals too. The quality of crops degrades due to harmful gases present in the environment. It has become fundamental need to screen the nature of air in encompassing. Observing of air contamination utilizing conventional strategies is very costly and tedious work. Sensor innovation has risen as a financially savvy, simple and quick technique for checking surrounding air quality. Arduino based air quality observing framework measures the nature of air utilizing Arduino and showcases it as PPM on LCD screen just as on Blynk mobile application through web availability. The ringer goes off if the air quality drops down from indicated PPM esteem. For estimating air quality, we utilized MQ135 sensor as it distinguishes the

majority of the destructive gases and can gauge their sum precisely. Cell phones are most regularly utilized gadget, applications on versatile are anything but difficult to utilize contrasted with web server. Subsequently constant update of nature of air can be observed whenever anyplace. This framework can for the most part be introduced in businesses and houses where destructive gases can be found and can give alert at whatever point as far as possible is crossed.

II. LITERATURE REVIEW

The disadvantages of the traditional system are their huge size, enormous weight and unprecedented cost. These lead to scanty organization of the checking stations.[6] So as to be powerful, the areas of the checking stations need cautious arrangement on the grounds that the air contamination circumstance in urban territories is profoundly identified with human exercises (for example development exercises) and area subordinate (e.g., the traffic stifle focuses have a lot of more awful air quality than normal).Due to this awkward foundation it was hard to control the air contamination or measure the nature of air. Arduino Based Air Pollution Monitoring System screens the Air Quality over a portable application using web and will trigger an alert when the air quality goes down past a particular level, infers when there are total of dangerous gases present perceptible all around like NH3, smoke, benzene, NOx, CO2. The system will show the air quality in PPM on the LCD and on the application which makes it simpler to utilize.
MQ135 sensor is utilized for observing Air Quality as it identifies most unsafe gases and can quantify their sum precisely. Right now, it can screen the contamination level from anyplace utilizing your Cell phones or portable. This framework can be introduced anyplace and can additionally trigger some gadget when contamination goes past some level, buzzer will make individuals aware when air quality gets perilous.

III. PROBLEM DEFINITION

Air contamination is the nearness of contaminants noticeable all around that meddle with human wellbeing andproduce other hurtful condition effects. Due to medical problems brought about via air toxins there is a genuine spotlight on air contamination checking exercises. Clean air is fundamental requirement for each individual. [4] Dirtied air causes numerous medical issues and a few harms. In this way to make any progression in front of controlling the contamination rate it is important to screen the air quality which may assist us with making a correct choice at perfect time. There are different reasons for expanding the contamination, for example, smoke vehicle exhaust, concoction release from ventures, radioactive substance etc. These are principle reason of diminishing the nature of air. The cardinal gases which truly impact the human prosperity are carbon monoxide (CO), hydrogen sulfide, sulfur dioxide (SO2), Nitrogen dioxide (NO2) and the standard duty of these gases are traffic related sullying release.

Tremendous endeavors are required to improve the nature of air in both outside and indoor condition. Observing of condition has been controlled from manual to the programmed control bit by bit. There is different improvement in the instrument of condition checking yet at the same time can't meet the unforgiving condition.

IV. PROPOSED WORK

All the hardware components are connected to each other with the help of jumper wires.[2] Arduino Uno is a microcontroller board dependent on ATmega328P. Arduino is simply connected to the computer via USB cable. Here computer can act as power source or an external power supply is given to the board. As soon as the power supply is provided to the system MQ135 gas sensor collects the data of gases present in the surrounding.MQ135 sensor is able to sense gases like ammonia, oxides of nitrogen, alcohol, benzene, smoke, carbon dioxide. This data from sensor is transferred to Arduino Uno in the form of voltage Level. Further these voltage levels are converted into PPM (parts per million) using a library for MQ135 gas sensor. The safe air quality level is 350 PPM and it should not exceed the limit of 1000PPM. If the value exceeds 1000 PPM it will result into various health issues such as heavy breathing, headaches, sleepiness, etc. When the sensor senses the value above 1000PPM the buzzer will be triggered. LCD screen will continuously display the quality of air in PPM as well as based on the value it will also display message-“Fresh air” if less than 1000PPM,”Bad air” if more than 1000PPM and “Danger” if more than 2000PPM.Parallely using Blynk mobile application the readings of air quality as well as message will be displayed on the respective mobile phone via internet. Here MQ135 sensor gives input to board, LCD and mobile application shows the output of the reading in PPM and buzzer will make people aware when air quality becomes dangerous. The way toward transferring yield on LCD or versatile application is finished with the assistance of Node MCU. It offers access to Wi-Fi or web. It can communicate with any microcontroller and it is the most driving gadgets in the IOT stage. Prior to starting the coding for this venture, we have to initially Calibrate the MQ135 Gas sensor. Here we are utilizing the Library for MQ135. Before this we have to adjust the MQ135 sensor, for aligning the sensor transfer the calibration code and let it run for 12 to 24 hours and afterward get the RZERO value While transferring the ensure that you

are associated with the Wi-Fi of your Node MCU gadget and the it will show yield on application.

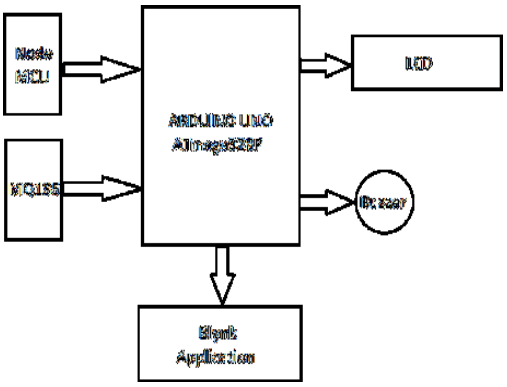


Figure 1. Block Diagram

V. METHODOLOGY

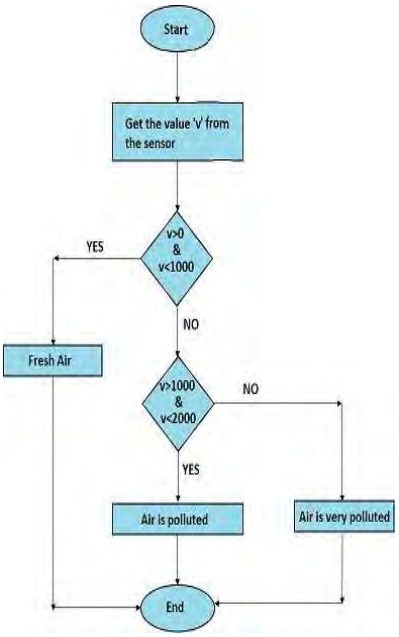


Figure 2. Flowchart

VI. RESULT & DISCUSSION

Arduino based air quality monitoring detector gadget design involves hardware and connection and eventually the collection of information from the detector via code for the Ardunio. To obtain actual time air pollution records (in PPM), we have taken reading using our customized sensor-detector in different environmental pollutions. We have used coil burning smoke, automobile smoke from street etc. The following values depicts for mosquito coil within 1 meter.

TABLE I. MOSQUITO COIL SMOKE DATA ANALYSIS

Time(second)	Value(PPM)	Distance (meter)
2	70(Fresh Air)	0.7
4	110(Polluted Air)	0.4
6	95(Fresh Air)	0.5

8	98(Fresh Air)	0.5
10	125(Polluted Air)	0.3
12	170(Polluted Air)	0.2
14	150(Polluted Air)	0.2
16	165(Polluted Air)	0.2
18	188(Polluted Air)	0.15
20	201(Very Polluted Air)	0.1

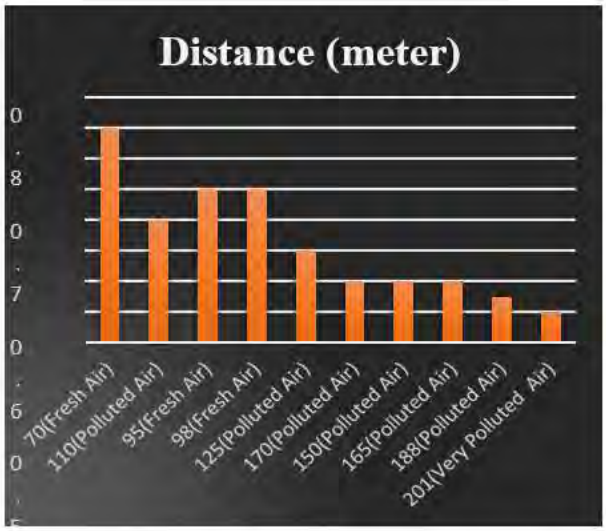


Figure 3. Graphical representation of data from coil smoke

From this facts analysis we can finish that for mosquito coil its miles safe to be a ways from the coil while burning at least 2 or three meters.

TABLE II. VEHICLE SMOKE DATA ANALYSIS

Time(second)	Value(PPM)	Distance (meter)
2	180(Polluted Air)	3.5
4	250 (Polluted Air)	2.5
6	270(Very Polluted Air)	2.3
8	245(Very Polluted Air)	2.6
10	260(Very Polluted Air)	2.4
12	200(Polluted Air)	3
14	290(Very Polluted Air)	2
16	230(Very Polluted Air)	2.7
18	199(Polluted Air)	3
20	215(Very Polluted Air)	2.8

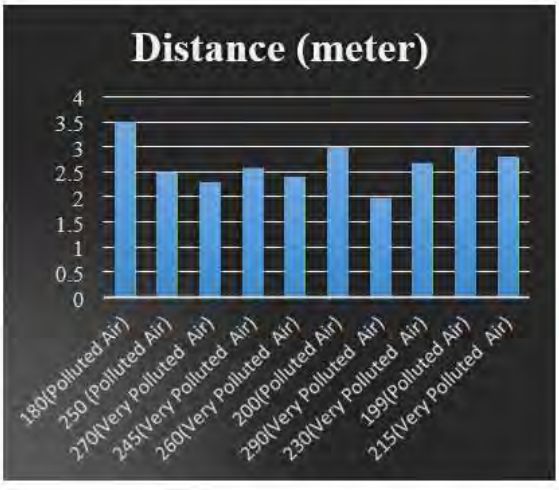


Figure 4. Graphical representation of data from vehicle smoke

From this facts analysis we can finish that for vehicle smoke its miles safe to be a ways from the smoke at least 4 or 5 meters.

VII. CONCLUSION

This task "Arduino based air quality checking framework" is skilled to identify the air quality given in a situation. It tracks the air quality and shows the given outcome as ppm and displays it on the LCD screen. In light of its presentation in the earth it tends to be said that it is anything but difficult to utilize and more affordable than the current systems. It is easy to understand just as effective to utilize. The framework can assists with making familiarity with the nature of air that one relaxes.

VIII. ACKNOWLEDGEMENT

We sincerely thank our guide Mr. Bhushan Nemade for his guidance and constant support and also for the stick to our backs. We likewise thank the undertaking organizers for masterminding the vital offices to complete the venture work. We thank the HOD, Dr. Rajesh Bansode, the Principal, Dr.B.
K. Mishra, and the school the executives for their help.

References

[1] Kim, Sunyoung, and Eric Paulos. "inAir: measuring and visualizing indoor air quality." Proceedings of the 11th international conference on Ubiquitous computing. ACM, 2009.

[2] Md. Abdullah Al Ahasan, Saumendu Roy, A. H. M. Saim, Rozina Akter, Md. Zakir Hossain,"Arduino-Based Real Time Air Quality and Pollution Monitoring System ". International Journal of Innovative Research in Computer Science & Technology (IJIRCST) Volume-6,
Issue-4, July 2018

[3] Krishna, V. Siva, and S. Arun. "Embedded System Based Air Pollution Detection in Vehicles." (2015).

[4] Muhammad Asif, Shafiullah Khan, Rashid Ahmad, Muhammad Sohail, Dhananjay Singh, "Quality of Service of Routing Protocols in Wireless Sensor Networks: A Review", Access IEEE, vol. 5, pp. 1846- 1871, 2017.

[5] J. S. Hwang, C. C. Chan, "Effects of air pollution on daily clinic visits for lower respiratory tract illness", American Journal of Epidemiology, vol. 155, no. 1,2002, pp. 1-10.

[6] P. Völgyesi, A. Nádas, X. Koutsoukos, A. Lédeczi, "Air Quality Monitoring with SensorMap", Proc of the 7th International Conference on Information Processing in Sensor Networks, pp. 529-

530, 2008.

[7] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, “Wireless sensor networks for habitat monitoring”, Proceedings of the 1st ACM International workshop on Wireless sensor networks and applications, Atlanta, Georgia, USA, 88-97, 2002.

[8] S. Choi, N. Kim, H. Cha, R. Ha, "Micro Sensor Node for Air Pollutant Monitoring: Hardware and Software Issues", Sensors, vol. 9, no. 10, pp. 7970-7987, 2009.

[9] Air_pollution_detector:https://en.wikipedia.org/wiki/Air_pollution_se nsor

[10] Arduino:https://www.arduino.cc/en/Main/ArduinoBoardUno https://en.wikipedia.org/wiki/Arduino

Smart Water Level Monitoring and Quality Monitoring System

Hemant Yadav, Jaikishan Yadav, Alok Yadav

Department of Information Technology Thakur College of Engineering & Technology, Mumbai, India
yhemant110@gmail.com, jaikiy15@gmail.com, yadavalok929@gmail.com

Abstract— **Internet of Thing from the past few years is emerging bent on be popular technology and it now features a huge demand in market thanks to its easy and easy integration with different technologies. Water may be a scarce resource and it should deplete over coming years thanks to overuse. The bad quality, overflowing water from tanks, leakage in pipes, and inefficient usage of water are the most cause which results in the wastage of water. So it's necessary have control on water wastage and usage furthermore by introducing or building a system which can overcome the water wastage related issues using Internet of Things (IOT) [6]. one amongst the main problems faced by most of the countries is that the issue of water scarcity within the world and wastage during transmission has been identified as a serious culprit; this is often one amongst the motivations for this research, to deploy computing techniques in creating a barrier to wastage so as to not only provide more financial gains and help the environment furthermore because the water cycle which successively ensures that we save water for our future. IOT based Water Level Monitoring system is an innovative system which can inform the users about the amount of liquid and can prevent it from overflowing. To demonstrate this the system makes use of containers, where the ultrasonic sensors placed over the containers to detect the liquid level and compare it with the container's depth. The system makes use of Arduino, GSM Module for sending data about the amount and quality to mobile phone and a buzzer. the ability supply is given through power jack in GSM and External battery in Arduino [1]. The liquid level is highlighted as distance from the underside in cms to point out the amount of liquid present within the container with the assistance of a SMS to the user. The buzzer starts ringing when the set limit of the liquid is crossed. Thus this method helps to forestall the wastage of water by informing about the liquid levels of the containers.Keywords – IOT, GSM, Arduino**

Introduction:

Water is one in all the natural resources that's most significant to our survival as humans, it's used for various purposes reception like for drinking, washing and bathing, in Agriculture it's used for irrigation, within the field of engineering it's employed in hydro power plants and atomic energy plants for power generation, it's also employed in the industries for various manufacturing needs. Its vast usage to our day to day activities highlights the necessity for correct management this natural resources. The common method of water level control for many homes is just to begin the feed pump at a coffee level and permit it to run until the next water level is reached within the cistern. This can be not properly supported for adequate system. Usually, this type of system provides visual multi-level also as continuous level indicator.

SMS notification at desired level and about the standard value supported user's requirements is included during this management system. Most of the buildings nowadays have overhead tank for water storage. Unfortunately, most users find it tedious and difficult to grasp the extent of water in their water tanks thanks to energy and time required to climb an overhead tank and therefore the absence of water level indicator. Thanks to this, there's wastage of huge volume of water. The answer to those problems is catered in our project, designing an automatic water level monitoring .This device includes a sensor which sense the extent of water within the cistern.

Problem Definition:

To design a microcontroller (IOT) based Water Level Monitoring that will continuously indicate the level of the water and automatically switch ON/OFF based on the threshold levels. The system should also measure the value of different quality parameter. Notification should be send through SMS about the water level and quality values and gives alert when water level reaches to threshold values.

Proposed System:

The proposed system is mainly developed to be implemented where water tanks are more commonly used such as Housing societies, Hotels, etc.

In this system, an application is developed which is used to indicate water level in the tank. The water levels are labeled as Low, Moderate and Full and gives Popup about the water level periodically [4]. The system also 87 sends notification about the quality of water through cell phone [3]. The Application also consist of controls through which the motor can be switched ON/OFF. The system provides the user with the feature to analyze the usage of water on monthly, weekly or daily basis from a website.

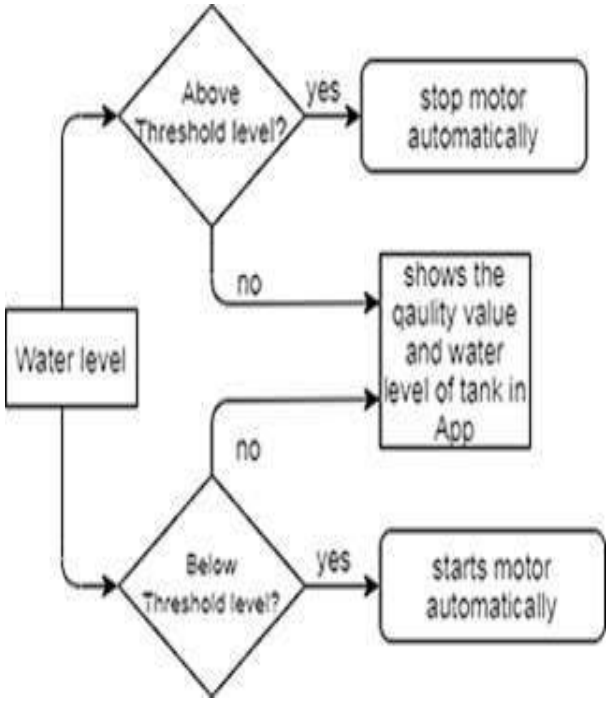


Fig 1: Flow diagram of working of system

Features :

→ **Water Maximization**: A water level controller can maximize the water usage and results in an appropriate level of water at all times being maintained.

→ **Live Notification**: User or the organization will get the live notification about the level of water and it's quality through SMS and also an indicator will be there indicating it's level through LED**.**

→ **Easy to Install:** Installation of all the systems are easy and the connection between them is quite simple for anyone.

Methodology

In this system we are going to check the water quality in the tank along with the water level. Water quality is also equally important. Due to some of the reasons water in the tank will be contaminated. For example dry leaves and dust. In the present system we initially connect the ultrasonic sensor to the core controller and place it on the top of the tank. pH sensor determines the amount of alkalinity present in the water pH range should always be between 0-14 for any liquid .The turbidity indicates the degree at which water loses its transparency. Relay and the water pump is used for pumping the water into the tank and it will stop automatically when tank is full. All the values of sensor and the level of water will be notified through SMS and for that we will use GSM Module. So there will be real time notification of the data to the user or organization. In this system we can control the flow of water in the tank using an application and also system can automatically switch off the flow of water only if tank is full. Hence by this system we can control the flow of water in tank manually as well as automatically [5]. User can monitor usage of water on daily, weekly, or monthly basis as the data is frequently uploaded at ThingSpeak which provides a platform to analyzing the data, visualize the data, etc. [2]

Feasibility study:

→**Technical Feasibility**: Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

→**Economic Feasibility**: As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

→**Operational Feasibility**: It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.

Understanding:

In this after understanding the topic the project feasibility was analyzed by performing different types of feasibility studies and by also planning the project tools, their project schedule, timeline charts, etc. Feasibility study will help in better understanding the various feasibilities associated with the project and helping to make the correct decisions and completing the project within the schedule, budget, etc.

The tools were specifically identified stating which technology can be feasible and how conveniently the project can be completed. This helps to understand the technology and tools that can be used for the project.

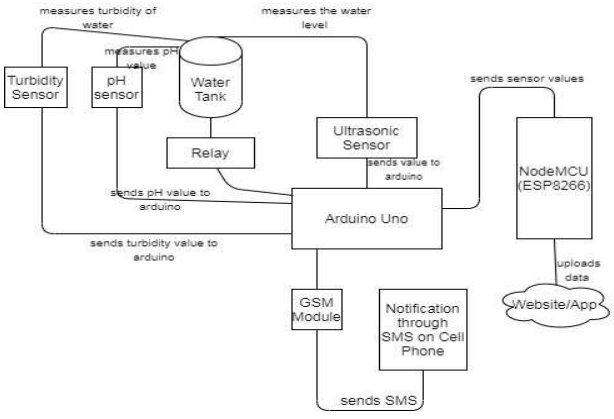


Fig 2: Block Diagram of System

Future Scope:

As the world's water resources become increasingly stressed, effective systems for management become more important. Several water monitor systems are available but most of them are either expensive or requires man power. Since wired technology is used in our proposed system there is scope to further modify it by using wireless RF technology. Thus, the communication between the controller and the driving element can be established wirelessly. Improvements can be made with minor changes in this model by eliminating the operator and providing the complete control to microcontroller (automatic level control). It can be used for level monitoring and control in industries. The system can also be extended to efficient functioning of dams. Therefore, a major future work can be possible in which a centralized control of all the dams in a state using GPRS or other wireless technology under central government can be beneficial to the whole country. On a local level, the control of all the water storage tanks in a society using wireless technology under a trusted authority can be beneficial as well.

I. CONCLUSION

The project is based on a single water level monitoring system with different levels indicated. The proposed mechanism of water control reduces the water wastage, ensures efficient use of available water resources and generates more precise and accurate results. There is no requirement of human laborer for monitoring the level, because the system manages the water level automatically. Also, operation execution time is less. Because of its cost efficiency this system can be installed in various rural areas where the water problems are on a rise. We have also included the GSM based system where the notification is send through SMS to the user. Also we have included the pH value and turbidity for the water quality analysis which helps us to improve the better understanding of the comparative efficiency of major drinking water disinfection practices against the currently most important resistant viruses.

References

[1] Joao Batista Rosolem ,Danilo Cesar Dini,Rivael strobel penze,Claudio Floridia,"Fiber Optic Bending Sensor for Water Level Monitoring",in IEEE Sensors Journal,(Volume: 13 , Issue: 11 , Nov. 2013)

[2] Lukas,Wisena Aditya Tanumihardja,Edy Gunawan," Monitoring of troughs water level using WSN",in 2015 IEEE Conference on Wireless Sensors (ICWiSe)

[3] James P.R. Sorensen, Andrew S. Butcher,"Water Level Monitoring Pressure Transducers",in British Geological Survey NERC. Ground Water Monitoring & Remediation

[4] Leonardo Alfonso, Arnold Lobbrecht,Roland Price," Information theory–based approach for location of monitoring water level gauges in polders,"in the American Geophysical Union.

[5] ChoZinMyint,Lenin Gopal,Yan Lin Aung,"Reconfigurable smart water quality monitoring system in iot environment", in 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)

[6] Mo Xiaocong,Qiu Xin Jiao,Shen Shaohong," An IoT based system for water quality monitoring",in 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics.

IOT based Intruder Alert System using Raspberry Pi, Pi camera and PIR sensor

Mr. Nikhil Valmiki, Mr. Shivam Tiwari, Mr. Amit Singh
Information Technology Thakur College of Engineering & Technology Mumbai
nikhilval1999@gmail.com, tiwarisomprakash.2000@gmail.com, singhamitrocks1999@gmail.com

Abstract— **In the realm of Internet Of Things (IOT) when we have all the innovations to alter our life, it's a good thought to build up a framework which can be controlled and observed from anyplace. There are numerous sorts of good security frameworks and cameras out there for home security yet they are a lot of costly so we will construct a minimal effort basic Raspberry Pi based Intruder Alert System, which alert you through an email as well as sends the image of Intruder when it recognizes any. This framework distinguishes the nearness of Intruder and rapidly alarms the house proprietor by sending an email alert. This mail will likewise contain the image of Intruder caught by Pi camera. This framework can be introduced at the fundamental entryway of office or home and can be checked from anyplace on the planet utilizing email over the web.**

Keywords – IOT, Raspberry pi, Pir sensors,security

I. INTRODUCTION

Remote security is the prevention of unauthorized access. The point of this task is to explore a practical arrangement that will give controlling of home apparatuses remotely and will likewise empower home protection from interruption without mortgage holder. The framework utilizes most recent remote correspondence like Bluetooth, Infrared and Wi-Fi access to the framework for security and mechanized machine control. Home security is viewed as a significant issue where wrongdoing is expanding and everyone needs to take appropriate measures to forestall interruption. SMS notice at wanted level and about the quality worth dependent on client's necessities can be remembered for this administration framework.

II. PROBLEM DEFINITION

- Getting Raspberry pi ready with its fundamental setup
- Getting pi-camera and Pir sensors ready
- Gathering the logic for the project
- Testing with data gathering and execution
- E-mail notification
- E-mail investigation with respect to proper alert and accuracy
    - Final deployment of the system

Every one of the previously mentioned themes has its own intricacy The configuration of the point is as per the following. It begins with the learning destinations, subject review, recommended readings and exercises that might be a few hands-on ventures followed by a speedy test to test the understanding, conversation themes for in the class and for email and afterward some valuable web joins giving most recent references identified with the point. All points follow a comparable example.

III. HARDWARE USED

The Raspberry Pi is a mini computer by the Raspberry Pi Foundation to advance the educating of essential software engineering in schools and in creating nations. The first model became unquestionably more well known than foreseen, selling outside its objective market for utilizations, for example, mechanical technology. It does exclude peripherals, (for example, consoles, mice and cases). In any case, a few adornments have been remembered for a few official and informal packs.

Infrared radiation is electromagnetic radiation (EMR) with longer wavelengths than those of observable light, and is thusly vague to the human eye. It is now and again called infrared light. The next task right now is to detached (PIR sensor) which is an electronic sensor that uses infrared (IR) light exuding from objects in its field of view. They are routinely used in PIR-based development discoverers. All articles with a temperature above preeminent zero exude heat essentialness as radiation. Commonly, this radiation is impalpable to the human eye since it transmits at infrared wavelengths, yet it might be distinguished by electronic devices proposed for such an explanation. The term uninvolved right now to the manner in which that PIR contraptions don't make or radiate any essentialness for area purposes. They work totally by recognizing the essentialness transmitted by various articles. PIR sensors don't perceive or measure "heat"; rather, they recognize the infrared radiation released or reflected from an item. The Raspberry Pi Camera Module v2 supplanted the first Camera Module in April 2016. The PI Camera Module has a 5-megapixel sensor Vision The Camera Module can be utilized to take top notch video, just as stills photos. It's anything but difficult to use for

fledglings, yet has bounty to offer propelled clients in case you're hoping to extend your insight. There are loads of models online of individuals utilizing it for time-pass, slow-movement, and other video shrewdness. You can likewise utilize the libraries we group with the camera to make impacts.

IV. WORKING

Framework will deal with the different distinctive remote correspondences and most recent mobiles for security reason. The proposed framework qualities include remote controlling of machines, interruption location, framework security and auto arrangement with the end goal that framework naturally modifies the framework settings on running equipment bolster check. The primary goals of the framework are: To create home security framework that

- Wrongdoing can be forestalled by making quick move.

- To the mortgage holder caught picture and sound will be appeared to interloper. Kids at home will stay safe because of approval and confirmation of a framework.

- Effective correspondence will happen by utilizing remote correspondence innovation.

Design and Architecture:

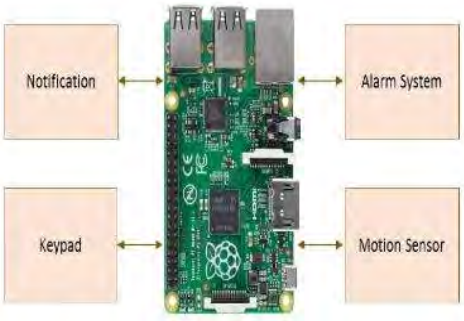


Fig .1. Block Diagram of IOT based Home Security

- Segments related with Home Security, (For example, PIR and piezo ringer and so forth.) are associated with the R-PI as portrayed in the figure above. The Movement sensor for example PIR sensor which represents Uninvolved Infrared Sensor is utilized to identify the nearness of gatecrasher in the home. Piezo Signal is utilized to raise an alert when the nearness of a gatecrasher is distinguished in the house with the assistance of PIR Movement sensor. Keypad is utilized to arm or incapacitate the security of the framework with the assistance of remarkable password for them two individually.

- The Framework contains different modules, for example, interruption identification, movement discovery and alert are examined in detail underneath:

Interruption Location: People emanate heat in type of radiation which isn't noticeable to the unaided eye, despite the fact that they can't be seen it very well may be distinguished. Latent Infrared Sensor recognizes this radiation to identify the nearness of any person. Hence PIR sensor is used to identify the nearness of gate crasher in the house.

- Alert: When the nearness of any human is distinguished through the PIR Movement Sensor and the present time is checked for, example in the event that the present time is past 11:00 PM, at that point the caution is raised.

- Warning: When the caution is raised a 30 seconds time period is designated to kill the alert, if the debilitate code is entered and on the off chance that it coordinates the correct impaired code got from the database inside the time frame(30 sec) at that point the caution is killed else the Proprietor and Law Implementations are told.

It expands the utilization of portable innovation to give basic security to our homes and for other control applications. The proposed home security framework catches data and transmits it to the individual mail utilizing the web. Raspberry pi works and controls movement locators and camcorders for remote detecting and reconnaissance, streams live video and records it for future playback. It can likewise locate the quantity of people situated with the assistance of the Infrared sensor. For instance, when movement is distinguished, the cameras naturally start recording and the Raspberry pi gadget alarms the proprietor of the conceivable interruption having an advanced cell. Raspberry-Pi has two primary parts communicating with one another. we can capable send it to the email consequently. We can ready to caution the individuals utilizing IOT innovation.

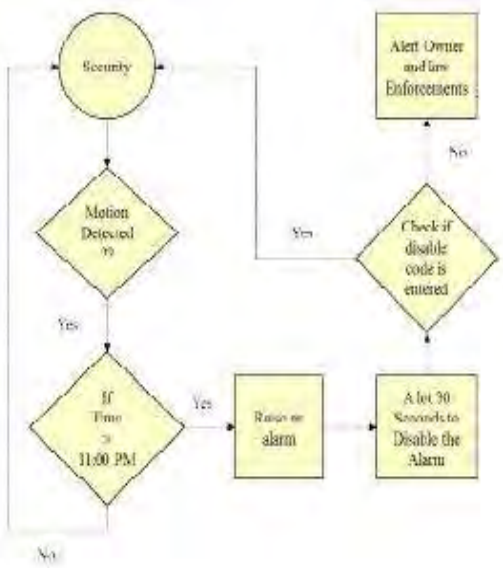


Fig.2.Flowchart of Home Security

V. RESULT & DISCUSSION

The surveillance controller system implemented using Raspberry Pi. Raspberry Pi is powered by 5V adapter. The desktop of Raspberry Pi is accessed remotely from putty software which is installed on laptop.

User seeks surveillance on webpage using Wi-Fi router's IP address. This system is useful for the owner to get a remote view of his home and to keep an eye on his valuables. If the face is detected, it is recognized and then camera is activated, image is captured and sent to the user. After that, motion software is triggered, image is recorded and notification is send to user.

Fig.3. Module of Home Security

VI. CONCLUSION

We have planned and built up an ongoing reconnaissance framework utilizing IOT module and Raspberry Pi. It is a functioning reconnaissance framework which will alarm the client when the occasion occurs. Live picture gushing is an extra bit of leeway of this framework. we have made web server which causes the client to see the live picture. This framework sends gatecrasher's caught video to the proprietor by the android versatile. The IOT based shrewd observation framework has been meant to plan so that it can satisfy the necessities of the client for specific reconnaissance region. It has incalculable applications and can be utilized in various situations and situations.

References

1. https://www.academia.edu/38573390/IoT_BASED_HOME_SECURITY_SYSTEM_USING_RASPBERRY_PI-3
2. https://circuitdigest.com/microcontroller-projects/raspberry-pi-iot-intruder-alert-system
3. https://pdfs.semanticscholar.org/658d/944982b3f098109f273cff2c2988e5b9b05a.pdf
4. https://www.irjet.net/archives/V5/i3/IRJET-V5I3740.pdf
5. http://www.ijircce.com/upload/2018/april/101_IOT%20Based%20Home%20Security%20System%20Using%20Raspberry%20PI.pdf

Fish Tank Monitoring System

By Jayesh Pandey, Dheeraj Poojary, Nigam Ojha
Thakur College of Engineering & Technology, TCET MUMBA

Abstract—**In light of the exploration, there are a portion of the comparable items identified with this present venture. One of the items which known as "A Multi- utilitarian Aquarium Equipped with Automated Thermal Control/Fodder-Feeding/water Treatment utilizing a Network Remote Control System" create by (Min-Chie Chiu, 2010). Subsequent to dissecting on their item, found that there have a couple of solidarity and shortcoming on their item. In the item, (PC) is utilized as primary controller to deal with all the sensors. The sensors were associated with the PC controlling framework through various module. The module is the ADC (Analog to Digital) which convert the simple sign from the sensors to advanced sign with the goal that the PC can peruse and group the qualities. At that point, by means of the VB interface, the customer PC can speak with server-PC to screen and control the aquarium dependent on the information from sensors. The quality of the item is the utilization of PC as controlling framework. PC has high preparing force and it can control the information got from the sensors by means of modules quicker contrasted with microcontroller which has constrained handling power. Be that as it may, there are a few shortcomings existing right now. From the start is the expense of the item. The item comprises of a PC as focal point of the framework which control and deal with the entire framework. It is extremely costly to distribute an entire PC just to deal with the aquarium framework. Second is the versatility of the item. The utilization of PC as focus some portion of framework make it hard to move the item starting with one spot then onto the next. The answer for the expense and convenientce issue can be tackled utilizing a microcontroller called Arduino Uno, to supplant the PC to work as principle controller of the framework to work the framework.**

Keywords:- IOT,ultrasonic sensor,Esp 8266

The term “Internet of Things (IoT)” (Huang & Li, 2010; Uckelmann, Harrison, & Michahelles, 2011) has been around for the past few years and is gaining recognition with the breakthrough of advanced wireless technology. Although there are several definitions for the phrase “Internet of Things” the best interpretation would be the integration of the physical world with the virtual world of the Internet. IoT is a network of physical objects that contain embedded technology to communicate and sense or interact with their internal or the external environment. Throughout the years, there have been many products built based on the IoT concept, for example, smartphones, smart watches, laptops, household electronic devices and other monitoring electronic system devices. Among these devices, the aquarium monitoring system is an electronic monitoring system that is grabbing a lot of attention nowadays.

What is an aquarium? An aquarium is a container or an artificial pond in which living aquatic animals or plants are kept. Aquarium is used for fishkeeping purpose for hobby and for indoor and outdoor decoration. Besides that, fishkeeping in aquarium is also related to various cultures, for example it is related to Vastu for Indians and Feng Shui for Chinese. According to Vastu, the aquarium is a perfect combination of harmony and balance. The water symbolizes the flow of life, growth and activities of living things. The motion and sounds of the bubbling water as it moves throughout the fish tank activates and increases the positive energy around the area, thus, bring good fortune, wealth and abundance. Meanwhile, according to Feng Shui principles, placing an aquarium in the home or office is an excellent way to attract auspicious chi into the space, especially for good luck, abundance and prosperity. The aquarium monitoring system was then introduced order to properly maintain the aquarium. Besides that, with the help of Iot devices such as aquarium monitoring system, users can monitor and manage their aquariums and fishes, anytime and from anywhere through the internet.

I. INTRODUCTION:

Presently the world is moving towards automation,everyone needs a solace living. In this day and age of connectedness, individuals are getting familiar with simple access to data. So by utilizing this innovation manual work is being diminished to incredible degree . In this manner, by this task we are endeavoring to supplant the manual observing of an aquarium with remote mechanization by means of web . In existing frameworks one can't sustain the fish in their nonappearance, it might bring theat to angles, temperature of the water can't be resolved as it very well may be changed by the seasons and fishes may experience the ill effects of this.It's difficult to screen everything going on inside your aquarium nonstop . With ordinary consideration and upkeep, you're probably going to get issues on the off chance that they happen just at your essence. Oceanic lives in the aquarium are effortlessly influenced with the progressions of the aquarium conditions, for example, temperature, bolstering, and the lighting. These conditions are basic for the getting by of the sea-going lives in the aquarium.

Right now world articles speak with one another and demonstration astutely. Web of Things is a developing innovation which understands this present nature of processing. Aquarium the executives needs opportune get-together of water parameter esteem changes. These progressions may influence the life of amphibian creatures in the aquarium. The proposed framework gathers sick supplant the manual upkeep of fish aquarium with its computerized capacities. the ongoing information from aquarium condition utilizing sensors, forms it and updates the temperature of water continuously accordingly of any

negative circumstances through temperature test. The undertaking's crowd is the gathering of individuals intrigued to keep angles at their homes or workplaces however don't have the opportunity to deal with, or they are stressed to continue soliciting their neighbors to take care from the fishes in their nonattendance. The venture is a computerized framework to deal with angles.

1.2 Problem Definition:

The time, labor, fish passing on because of deficient upkeep of aquarium is more. To beat these issues we got ready for shrewd aquarium to tackle all of referenced above issues.

II. PERIODS OF PROJECT

● Phase 1: Planning, Analysis, Designing and Implementation.

Examination: Getting away from of the venture title and doing research on it we will get our definition and after that then we will initially make the Literature Survey of the extend and do the entire documentation.

Arranging: After investigation we will initially learn about it and do some examination on it for our better comprehension of the undertaking and furthermore getting an unpleasant picture about what might be our concern definition for the specific venture.

Structuring: Then we will build the plan of the undertaking and as per that, will list down all the necessities required for the development for the model of our task.

Usage: After securing the necessities we initially built up the associations of the equipment segments (Arduino, Ultrasonic sensor and so on) and afterward included the coding part later with assistance of Arduino information link to the Arduino.

Phase 2: Testing and Deployment.

Testing: After the model is prepared we will initially interface the equipment with the doled out code and afterward we will check on the off chance that it bolsters the instrument or not. If not we will explain the issues with respect to it and will check once more.

Arrangement: After, complete mix and testing of task constant running and activity of the framework will be finished and useful.

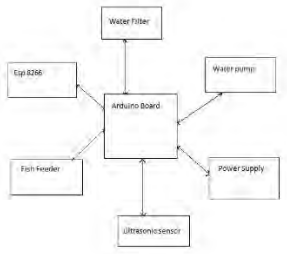
III. METHODOLOGY USED:

The square graph of Arduino based aquarium observing framework is appeared in figure1 Arduino board are use as the primary controller to screen and control the sensors,lights,radiator. The Arduino Uno is essentially a little microcontroller. Here the force transfer is utilized to control the aquarium lights, the dc engine is utilized to angle nourishing reason, temperature sensor is utilized to detect the water temperature and sends the warning in like clockwork, if the water temperature is low the warmer is utilized to changes the ordinary temperature. The LCD show is utilized to show the temperature and by observing LCD show.We can work the lights and dc engine for nourishing fishes and This undertaking can be work in three distinct modes. Web Of Things (IOT), it sends refreshes in at regular intervals so we can screen the aquarium temperatures when we are away from home.Utilizing infrared sensors it can worked by remote control On-Screen menu that can be constrained by up-down catches by observing the LCD show.

2.2 Summary:

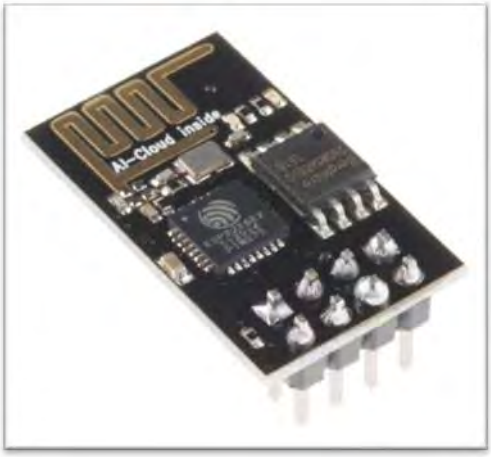
This audit is proposed to help a total computerized aquarium . Microcontroller in the framework guarantees about increment in framework's life by lessening the force utilization bringing about lower power utilization. Computerized aquarium framework has a tremendous interest and future degree in fish cultivating for large scale manufacturing of fishes. It is efficient, prompted evacuation of human blunder in modifying accessible water levels and to amplify their net benefits in understanding to factors like deals, quality and development of their item.

2.3 Block diagram



2.4 System Components: ESP8266:

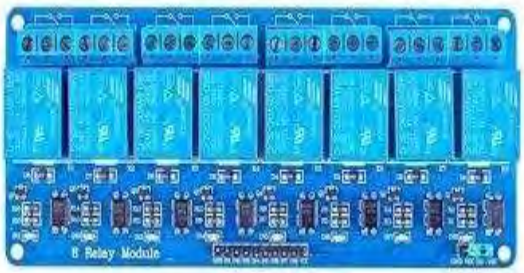
The ESP8266 is an ease Wi-Fi microchip with full TCP/IP stack and microcontroller capacity delivered by maker Espressif Systems in Shanghai, China. The chip previously went to the consideration of Western producers in August 2014 with the ESP-01 module, made by an outsider maker Ai-Thinker.



Ultrasonic sensor:

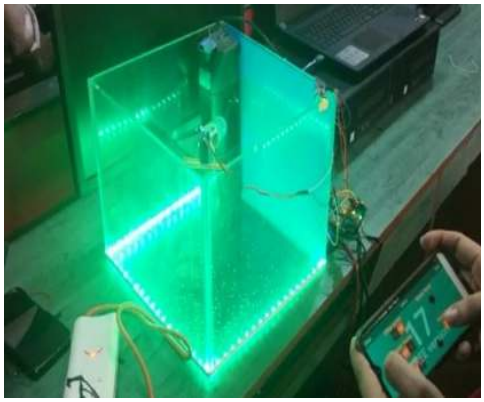
Ultrasonic sensors work by radiating sound waves at a recurrence unreasonably high for people to hear. They at that point trust that the sound will be reflected back, figuring separation dependent on the time required. This is like how radar quantifies the time it takes a radio wave to return in the wake of hitting an article.

8-relay channel:



This is a 8-channel LOW level trigger relay and it very well may be applied to Arduino and Raspberry Pi. Relays are reasonable for driving high force electronic gadgets, for example, lights, electric fans and cool. A relay can be utilized to control high voltages with a low voltage by interfacing it to a MCU.

IV. RESULT:



V. CONCLUSION:

In the present world everything is moving towards motorization to diminish work weight of human. As indicated by the composing overview as of late arranged systems over not too capable at their working and moreover not financially reliable in the point of view on use. Particularly microcontroller based introduced systems has less memory size close by the most distant point in interfacing peripherals this prompts slack in the plan of the structures. In like manner here proposed and automated Arduino based aquarium watching structure that engages straightforward direction over the couple of issues of aquarium, for instance, temperature assortments, empowering timetable, turbidity level, lighting system which contains sensors to aggregate the information required for needed incitation using Arduino controller. Which gives beneficial electronic direction over the issues.

VI. FUTURE SCOPE:

The rule purpose of this endeavor is to develop a structure that can screen the status of the aquarium reliably and engage the customer to screen status of the aquarium by marking on to the database system and perform appropriate exercises through the web itself or let the structure itself perform instructed decisions and exercises. Other than that, system furthermore should be flexible, accompanied straightforwardness, uses low power, supportive and successful. To vanquish the issues referenced more than, a couple of measures have been executed for this endeavor:

The incorporation of this system is obliged to little quantify aquariums, for instance, 10-30 galloon fish aquarium.

The point of confinement of the region used is simply inside a particular extent of the home or office which has a not too bad web incorporation, as the system needs to

invigorate the customer on the status of the aquarium by methods for the web.
The adventure requires programming capacity to manage the Arduino uno.

References

[1]. fish talk: an iot based mini aquarium system yi-bing lin ; hung-chun tseng ieee access. year:2018, volume:7, pages:35457-35469. ieee journals &magazines.

[2]. a novel design and implementation of automated feeding mechanism in fish aquarium mohammad abdul hye ; md manjurul akter ; atiq mohammad jahangir ; hasan u. zaman 2018 2nd international conference on electronics, materials engineering and nano- technology (iementech). year: 2018, ieee conferences.

[3]. automated aquaculture system that regulates ph, temperature and ammonia. aaron don m. africa; jeremy czar christian a. aguilar; charles martin s. lim; paulo arnel a. pacheco; steven edward c. rodrin,

[4]. 2017 ieee 9th international conference on humanoid,

[5]. nanotechnology, information technology, communication and control, environment and management(hnicem). year:2017, ieee conferences.

[6]. nurliani hidayah ritonga; agung nugroho jati; rifki wijaya, 2016, may. automatic arowana raiser controller using mobile application based on android. in 2016 ieee asia pacific conference on wireless and mobile (apwimob) (pp.86-87). ieee.

[7]. prasad, a.n., mamun, k.a., islam, f.r. and haqva, h., 2015, december. smart water quality monitoring system. in

[8]. 2015 2nd asia-pacific world congress on computer

[9]. science and engineering (apwc on cse) (pp. 1-6). ieee.

[10]. automatic aquarium monitoring for the fish farm aquaculture environment jui-ho chen ; wen-tsai sung ; guo-yan lin 2015 ieee international conference on systems, man and cybernetics. year: 2015, ieee conferences

[11]. taotao xu , feng chen , 2014 august. an embedded fuzzy decision system for aquaculture. in 2014 2014 ieee workshop on electronics, computer and applications. (pp.351-353). ieee conference publications.

[12]. noor, m.z.h., hussian, a.k., saaid, m.f., ali, m.s.a.m. and zolkapli, m., 2012, july. the design and development of automatic fish feeder system using pic microcontroller. in control and system graduate research colloquium (icsgrc), 2012 ieee (pp. 343- 347). ieee. multi-functional aquarium equipped with automatic thermal control/fodder- feeding/water treatment using a network remote control system. information technology journal, 9(7), pp.1458-1466.

Smart Street Light System

Nikhil Vishwakarma, Maitri Vora, Akshaya More
TCET, MUMBAI
menikhilv143@gmail.com, voramaitry25@gmail.com, akshayamore37@gmail.com

Abstract— **Streetlights are the basic a part of any city since it facilitates better night visions, secure roads, and exposure to public areas but it consumes a quite large proportion of electricity. within the manual streetlight system lights its powered on from sunset to sunrise with maximum intensity even when there's sufficient light available. This energy wastage may be avoided by switching off lights automatically. The saved energy may be efficiently utilized for other purposes like residential, commercial, transportation etc. this will be achieved using an IOT enabled streetlight management system. The project uses Light Emitting Diodes (LED) that don't consume an infinite amount of electricity to exchange the facility consuming traditional HID lamps. LED lights together with LDR enables the intensity variation which is infeasible with the HID lamps. As LEDs are directional light sources it can emit light in specific direction thereby optimizing the efficiency of the streetlights. This work is implemented employing a programmed Arduino board for providing the desired intensity of sunshine at various times. The proposed work has achieved a far better performance compared to the present system.**

Keywords – Street Lights; IOT; Light Emitting Diodes HID (High Intensity Discharge Lamps); Arduino;Automatically.

I. INTRODUCTION

IOT is that the network of physical devices Now a days, human has become too busy, and is unable to search out time to change of the lights wherever not necessary. Streetlights are an integral a part of any developing locality. they're present on all major road-ways and within the suburbs too. this system is just like the lights are switched on within the evening before the sun sets and that they are transitioned the following day morning after there's sufficient light on the surface. the road lighting is one in all the biggest energy expenses for a city. An intelligent street lighting system can cut municipal street lighting costs the maximum amount as 50% - 70%. An intelligent street lighting system could be a system that adjusts light output supported usage and occupancy, i.e., 0automating classification of pedestrian versus cyclist, versus automotive. An intelligent street light management proposes the installation of the wireless based system to remotely track and control the particular energy consumption of the road lights and take appropriate energy consumption reduction measures through power conditioning and control. The smart street light controller must be installed on the sunshine pole which carries with it microcontroller together with various sensor and wireless module. The smart street light controller installed on the road light pole will control LED street lighting looking on movements of the thing within the street. Whenever a vehicle passes by, the LED will increase its intensity then decrease it whenever the roads are idle.

II. PROBLEM STATEMENT

In most of the cities, the road lights are ON when it's not need and it's OFF when isn't needed. due to these situation the large energy expenses for a city gets wasted. Usually the lights are ON within the evening after the sunset, it continuous to air till the sun rises within the subsequent day morning. This paper focuses on reducing the energy by automatically switching ON and OFF street lights. When vehicles come to the street/road the sensor will capture ande movements of the vehicles then lights automatically ON. Otherwise automatically OFF the lights. In sunny and rainy days, ON time and OFF time differ significantly which is one of the major disadvantage of using timer circuits or manual .

II. BACKGROUND

The key indicators of India's energy problems include; Over 40 per cent of the households (particularly rural areas) in India still don't have electricity, a couple of third of our total primary energy supply to rural areas still comes from non-commercial sources (biomass, dung) and currently India faces an infinite demand supply gap of about 15-25% energy shortage.

Due to shortage of the energy supply till today several villages haven't facilitated with electricity and whether or not provided, the provision of the electricity is proscribed to few hours in a very day and face serious problems because of unlimited power cuts. During the day time we get enormous amount of sunshine energy from sun and also the problem for pedestals are common during the evening time. Though most of the streets are equipped with street lights in each and each village areas but because of the uncontrolled power failures/power cut it's becoming a significant problem.

A simple and effective solution to the present problem would be dimming the lights during off peak hours. Whenever presence is detected, the lights around it'll glow at the conventional (bright) mode. this could save lots of energy and also reduce cost of operation of the streetlights. we are able to check the status of street light on internet using IOT (Internet of 01things) from anywhere in real time and solve the problems if happen during the processing.

III. IMPORTANCE OF THE PROJECT

The internet of things (IoT) are ready to implement transparently a really great deal of heterogeneous end systems, while digital service provides open access to sub

set of information. With Smart street light system it's possible to systematically help the pedestrian to succeed in the destination within the remote rural areas which face serious power supply problem.

• This project is targeted on the controlling intensity of the sunshine considering the item movement.

• This smart light system automatically detects the movements of the item on the road.

• This Smart light system is employed to scale back energy consumption.

• This smart system is employed to avoid unnecessary usage of electricity. the whole smart system is meant to control using artificial energy source.

• This smart system is best suited to street lighting in remote urban and rural areas where the traffic is extremely low.

IV. OBJECTIVES AND SCOPE

The main objective of this project is to implement a IoT based Automatic Street Light System. because the traffic decreases slowly during late-night hours, the intensity gets reduced progressively by detecting the movement of vehicles till morning to save lots of energy and thus, the road lights activate at the dusk then flip at the dawn, automatically. the method repeats a day..

White Light Emitting Diodes (LED) replaces conventional HID lamps in street lighting system to incorporate dimming feature. The intensity isn't possible to be controlled by the high intensity discharge (HID) lamp which is usually utilized in urban street lights. LED lights are the long run of lighting thanks to their low energy consumption and long life. LED lights are fast replacing conventional lights because intensity control is feasible by the heartbeat width modulation.

This proposed system uses an Arduino board. Strings of LED are interfaced to the Arduino board. A programmed Arduino board is engaged to produce different intensities at different times of the night in keeping with the movement of vehicles. This project is enhanced by integrating the LDR to follow the switching operation precisely and IOT to display the status of street on browser and help in controlling it.
The main objectives are as follows:

• To avoid unnecessary Waste of sunshine.

• Provide efficient, automatic and smart lightning system.

• Totally supported Renewable energy sources.

• Longer life.

• Energy Saving

VI. PHASES

- Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: Getting clear idea of the project title and doing research on that we are going to get our definition and at the moment then we are going to first create the Literature Survey of the project and do the entire documentation.

Planning: After analysis we are going to first study about it and do some research on that for our better understanding of the project and also getting a rough picture about what would be our problem definition for the actual project.

Designing: Then we are going to construct the look of the project and in line with that, will list down all the necessities needed for the development for the prototype of our project.

Implementation: After acquiring the necessities we first developed the connections of the hardware components (Arduino, MQ2 sensor etc) so added the coding part later with help of Arduino data cable to the Arduino.

- Phase 2: Testing and Deployment.

Testing: After the prototype is prepared we are going to first connect the hardware with the assigned code and so we are going to check if it supports the mechanism or not. If not we are going to solve the problems regarding to that and can check again.

Deployment: After, complete integration and testing of project real time running and operation of the system are done. User must check the worth.

VII. METHODOLOGY

Recent days, Smart Street Light System is major component of a sensible City Infrastructure. The important function is lighting town streets using sensor's to avoid wasting the present or power energy
.In existing system using normal street lamps. It takes more current and costs too. So use LED lamps to avoid wasting the present in low amount of power.

VIII. HARDWARE

1. ARDUINO UNO: Arduino may be a single-board microcontroller, intended to form the applying of interactive objects or environments more accessible. The hardware consists of an open-source hardware board

designed around an 8-bit Atmel AVR microcontroller, or a 32-bit Atmel ARM. Current models feature a USB interface, 6 analog input pins, yet as 14 digital I/O pins which allows the user to connect various extension boards. The Arduino board exposes most of the microcontroller's I/O pins to be used by other circuits. The Diecimila, Duemilanove, and current Uno provide 14 digital I/O pins, six of which may produce pulse width modulated signals, and 6 analog inputs.

2. IR SENSORS: An infrared sensor is an instrument that's accustomed sense certain characteristics of its surroundings by either emitting and/or detecting actinic ray. it's also capable of measuring heat of an object and detecting motion. Infrared waves don't seem to be visible to the human eye. within the spectrum, actinic ray is that the region having wavelengths longer than actinic radiation wavelengths, but shorter than microwaves

Key benefits of infrared sensors include low power requirements, simple circuitry and their portable feature. the primary IR LED is wired to emit IR waves and therefore the second LED is wired to transmit a proof when it receives an IR input. When an object comes within range of the emitting LED, it reflects the IR back to the receiving LED and produces a proof.

3. LIGHT DEPENDENT RESISTORS:

A LDR or a photograph resistor may be a device whose resistivity may be a function of the incident radiation. Hence, they're light sensitive devices. they're also called as photo conductors, photo conductive cells or just photocells. When light falls i.e. when the photons fall on the device, the electrons within the valence band of the semiconductor material are excited to the

conduction band. These photons within the incident light should have energy greater than the band gap of the semiconductor material to form the electrons jump from the valence band to the conduction band.

Hence when light having enough energy is incident on the device more & more electrons are excited to the conduction band which ends up in sizable amount of charge carriers.

4. LED: An LED may be a two-lead semiconductor source of illumination, which emits lights when activated. When an appropriate voltage is applied to the LED terminal, then the electrons are able to recombine with the electron holes within the device and release energy within the sort of photons. the colour of the LED is decided by the energy band gap of the semiconductor. The forward voltage require to show ON a LED, depends on the colour of the LED. If you're feeding the precise value of forward voltage then you'll be able to connect a LED on to the source. If the voltage may be a on top of use a resistance serial with the LED.

5. JUMPER WIRES: Jumper wires are used for creating connections between items on your breadboard and your Arduino's header pins.

6. USB CABLE: this can be a regular A-B USB Cable. It will be accustomed connect your computer to Arduinos that use a full-sized B-type USB connection, like the Uno and Mega2560.

7. BREADBOARD: Breadboard provides temporary connections between individual electronic components, to style and assemble demonstration circuits. Breadboard is convenient because you'll be able to quickly change the circuit without excessive or complicated modifications, making breadboard ideal for circuit development and testing. Breadboard consists of a grid of contacts with contacts typically
spaced one-tenth of an in. apart.

IX. BLOCK DIAGRAM

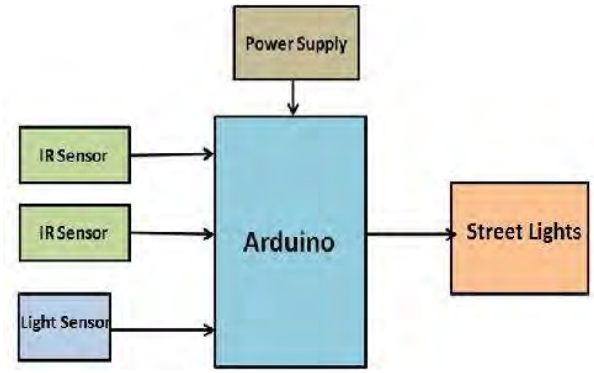


FIG1: BLOCK DIAGRAM

X. CIRCUIT DIAGRAM

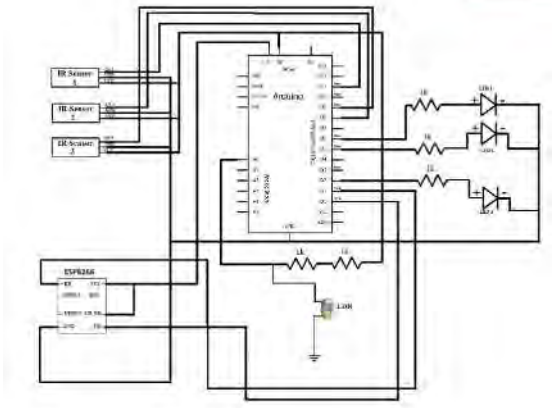


FIG2: CIRCUIT DIAGRAM

XI. RESEARCH GAPS

Every project has some gaps in it which can be implemented in future. No project is 100% efficient. There are some gaps in this project too. They are as follows:
1. As this system is currently not WiFi enabled, we cannot detect which LEDs are not working. By adding WiFi

module we can detect damaged LEDs from a web server and the manual help can be sent only to those locations.
2. Also, the system does not GPS connection so we cannot detect the exact location of LED.
3. A SOS button can also be implemented in future, which can send notification whenever some emergency in the surrounding is detected.

XII. CONCLUSION

Uuse of power electronics is increasing exponentially across various sectors of human life. The components employed in the project, like Arduino and sensors are slowly becoming an important a part of our daily routines. So, it's only fitting that we use them to boost efficiency in every walk of life. Keeping in mind the urgent need for energy conservation, Smart Street Light System with IoT is a wonderful and effective solution. It combines safe lighting protocols with consumption of minimal amount of power by reducing the intensity of the lights when not in use whereas in present system the lights are on fully intensity even when the road is totally empty which causes unnecessary use of electricity..

References

[1] “Design of Multi-functional Street Light Control System Based on AT89S52 Single-chip Microcomputer” Hengyu Wu, Minli Tang, IEEE, 2010.
[2] “A New Streetlight Monitoring System Based On Wireless Sensor Networks” IEEE 2010 [
3] “Automatic Street Light Intensity Control and Road Safety Module Using Embedded System International Conference on Computing and Control Engineering (ICCCE 2012), 12 & 13 April, 2012.
[4] “Intelligent Street Lighting System Using Gsm” International Journal of Engineering Science Invention ISSN
[5] “Analysis Of Solar Energy Based Street Light With Auto Tracking System”, International journal1,2012
[6] “Wireless internet lighting control system”,Budike, E.S. Lothar (Power web Technologies), US patent 7,167,777, Jan 23, 2007.
[7] http://www.ehow.com/howdoes_5561845_do-is ensorswork.html#ixzz310syLD6I
[8] http://homepages.which.net/~paul.hills/Emc/Be cBody.html
[9] http://en.wikipedia.org/wiki/Arduino [10] http://www.galco.com/comp/prod/sensor.ht m
[11] http://www.passmyexams.co.uk/GCSE/phy sics/LM324.html

An Obstacle Avoiding Robot using Arduino

Tanvi Pradhan, Pradyumn Rai, Yash Sarda
TCET, Mumbai
pradhan.tanvi500@gmail.com, raipradyumn1999@gmail.com, yashsarda189@gmail.com

Abstract—**Trajectory planning is one of the most significant essential focuses in pick and spot undertakings done by mechanical controllers. Right now, we have displayed a robot, which is minimized, independent and completely practical. This robot or a brilliant vehicle is worked to detect any hindrance in its way, to keep away from it and helps in maintaining a strategic distance from the mishaps (accidents) to happen. IR sensors were adjusted to execute constant obstacle occurring framework for wheeled robots, with the goal that the robot can persistently identify environment, maintain a strategic distance from snags, and move securely. Right now obstruction avoider framework is being presented with snag evasion and furthermore mishap anticipation for vehicles. This will give a higher likelihood to diminish the mishaps occurring each day on the streets (particularly mishaps happening on the streets in uneven zones). Here an Arduino based model has been produced for mishap evasion which is ease and proficient. All through the development of this model, we instructed ourselves to the Arduino Coding language, and thoroughly with the working of an Infrared Sensor and its highlights. Taking everything into account, through this venture, we plan to develop a model of a savvy vehicle that is useful to the quotidian issues of the present age.**

Keywords – Arduino UNO, Motor Driver, IR Sensor, Servo Motor, Battery, Connectimg Wires, etc.

I. OVERVIEW

Obstacle avoidance is one of the most significant parts of versatile mechanical autonomy. Without it robot development would be exceptionally prohibitive and delicate. This report clarifies obstruction shirking utilizing infrared sensors. Presently day's numerous enterprises are utilizing robots because of their elevated level of execution and unwavering quality and which is an incredible assistance for people. The obstruction shirking mechanical autonomy is utilized for distinguishing snags and maintaining a strategic distance from the crash. This is a self-governing robot. The plan of obstruction evasion robot requires the reconciliation of numerous sensors as per their undertaking.

A. Background

From its introduction during the 1950s, present day robots have made some amazing progress and established itself as a permanent guide in the headway of mankind. Over the span of time, robots took numerous structures, in view of its application, and its size shifted from a mammoth 51 feet to infinitesimal level. Over the span of innovative improvements of robots, one angle stayed instrumental to their capacity, and that is versatility. The expression "obstruction evasion" is presently utilized in current mechanical technology to signify the capacity of robot to explore over an obscure domain without having any crash with encompassing articles (Arduino-Apply autonomy, 2013). Deterrent evasion in robots

can get greater adaptability moving in fluctuating situations and would be substantially more productive as ceaseless human observing isn't required. This venture built up a snag staying away from robot which can move with no crash by detecting obstructions on its course with the assistance of ultrasonic separation sensors. Robots guided with this innovation can be placed into broadened utilizes, e.g., looking over scenes, driver less vehicles, self-governing cleaning, computerized grass cutter and overseeing robot in ventures.

B. Importance of Project

- Use new technology that uses Internet of Things (IOT).
- To improve traversing in an unknown environment.
- To achieve cost savings.
- To enhance the ability of a device to research.
- Helps in damage control for a robot by avoiding obstacle.

C. Motivation

Numerous robots for mechanization and route have been created as of late like divider following, edge-following, human after and obstruction maintaining a strategic distance from robots. The obstruction keeping away from robot will avoid deterrents it experiences in its way towards its operational objective. Because of the unwavering quality, openness and cost viability of utilizing versatile robot in industry and specialized applications, the hindrance maintaining a strategic distance from robots are significant in manufacturing plant floor. Then again, Unmanned Aeronautical Vehicles (UAVs) are assuming a fundamental job in barrier just as non-military personnel applications. The military applications incorporate observation, reconnaissance, fight harm appraisal and correspondences. Then, non-military personnel applications incorporate debacle the board, remote detecting, and traffic observing, and so on. Huge numbers of the UAVs applications need the capacity to explore in urban condition or obscure territories that have numerous snags of various sorts and sizes. Fundamental necessity of self-governing UAVs is to distinguish snags in its way and stay away from them. This paper proposed

a case of the snag evading robot calculation and plan of the robot base utilizing IR and PIR sensors. The created robot can be utilized as a stage for a few applications in instructive, look into or mechanical.

D. Objective And Scope

The Objective of this project is to implement a low cost, reliable and scalable obstacle avoiding car that can be used in unknown terrain and carry out research work if needed. The Scope can be was as by adding new sensors we can advance the features by expanding and improvising according to need.

- To make a System easy and to avoid accidents.

- To help the aged / old ones to make their driving safer.

- These new technologies are not only helpful but also provide us with security while driving. The objective is to make driving safe using this IOT based technology.

II. LITERATURE SURVEY AND PROPOSED WORK

In this section, we present literature survey done related to this project and relating work in progress.

A. Introduction

Obstacle avoidance is an essential necessity of any self-ruling portable robot. Obstacle shirking robot is configuration to permit robot to explore in obscure condition by keeping away from crashes. Snag maintaining a strategic distance from robot detects obstructions in the way, keep away from it and resumes its running. There are some renowned strategies for robot route like divider following, edge discovery, line following. One of the business frameworks utilizes divider following technique on a story cleaning robot for long passages. A progressively broad and normally utilized technique for obstruction shirking depends tense identification. A burden with hindrance shirking dependent nervous distinguishing is the need of the robot to stop before a deterrent so as to give an increasingly exact estimation. Every single portable robot highlights a crash shirking, extending from crude calculations that identify an impediment and stop the robot so as to dodge an impact, utilizing some modern calculations that empower the robot to reroute snags. The last calculations are increasingly perplexing, since they include location of a deterrent just as a quantitative estimation concerning the obstruction's measurements. Once these have been resolved, the impediment shirking calculation needs to direct the robot around the snag and resume movement toward the first objective. Right now, directing calculation guarantees that the robot doesn't need to stop before a deterrent during its route. Thus, the robots may beat a portion of the issues during route, which are examined above and it can explore easily during its activity maintaining a strategic distance from the impacts. We have introduced an essential calculation and structure which can be additionally improved relying on the necessary applications.

B. Literature Survey

From this research paper we got definite data of the subject IoT. We got a thought of forestalling accidents with the assistance of IoT utilizing various segments. We got information about the pace of accidents occurring in India and how to forestall it and get decrement in the proportion of mishaps [1]. A Basic yet powerful snag avoider for the IARA Self- governing vehicle this paper gave a thought of utilizing IR Sensor for utilizing hindrance shirking [2].

C. Problem Definition

These days, mechanical advances have gotten increasingly significant since a great deal of industry is attempting to improve their hardware weapons. This innovation has created step by step to ensure a great outcome. As of late, by time passes by, a ton of mechanical robots have been designed to

help people groups running their day by day life. Snags Shirking Portable Robot is really a straightforward impact evasion machine. Other than that, its future advancement is extremely large to investigate. By utilizing this straightforward impact evasion framework, a ton of new and assortment versatile robot with numerous capacities can be developed. Constant obstruction evasion is one of the key issues to fruitful uses of versatile robot frameworks. Every single portable robot highlights a crash evasion, running from crude calculations that recognize a deterrent and steer the robot shy of it so as to dodge an impact, through complex calculations, that empower the robot to bypass hindrance. Once these have been resolved, the hindrance evasion calculation needs to control the robot around they obstruction and resume movement toward the first objective.

D. Phases of Project

- Phase 1: Arranging, Examination, Planning and Execution.

Arranging: After examination we learned about our subject and did some exploration on it for our better comprehension of the undertaking and furthermore to get an unpleasant picture about what might be our concern definition for the specific venture.

Examination: We got the away from of our task title by doing research on it and will conclude our definition and after that we previously made the Writing Overview of the venture and did the entire documentation.

Structuring: At that point we will develop the plan of the task and as per that, will list down all the necessities required for the development for the model of our venture.

Execution: In the wake of obtaining the necessities we continued for the development of the application.

- Phase 2: Testing and Arrangement.

Testing: After the model is prepared, we will initially interface the equipment with the allocated code and afterward we will check in the event that it underpins the instrument or not. If not, we will illuminate the issues with respect to it and will check once more.

Arrangement: After, complete coordination and testing of task ongoing running and activity of the framework will be finished. We are relied upon to approve against their recently put away information.

III. ANALYSIS AND PLANNING

Obstruction shirking is an essential necessity of any self-sufficient versatile robot. Snag evasion Robot is configuration to permit robot to explore in obscure condition by staying away from crashes. Hindrance maintaining a strategic distance from robot detects impediments in the way, dodge it and resumes its running. There are some well-known techniques for robot route like divider following, edge identification, line following. One of the business frameworks utilizes divider following strategy on a story cleaning robot for long foyers. A progressively broad and usually utilized strategy for snag evasion depends tense recognition. A
hindrance with deterrent evasion dependent nervous recognizing is the need of the robot to stop before a snag so as to give a progressively exact estimation. Every single portable robot highlights an impact evasion, running from crude calculations that identify a hindrance and stop the robot so as to maintain a strategic distance from a crash, utilizing some refined calculations that empower the robot to bypass impediments. The last calculations are progressively perplexing, since they include discovery of a snag just as a quantitative estimation concerning the deterrent's measurements. Once these have been resolved, the snag shirking calculation needs to control the robot around the impediment and resume movement toward the first objective. Right now, guiding calculation guarantees that the robot doesn't need to stop before an impediment during its route. Consequently, the robots may conquer a portion of the issues during route, which are talked about above and it can explore easily during its activity dodging the impacts. We have exhibited an essential calculation and plan which can be additionally improved relying on the necessary applications.

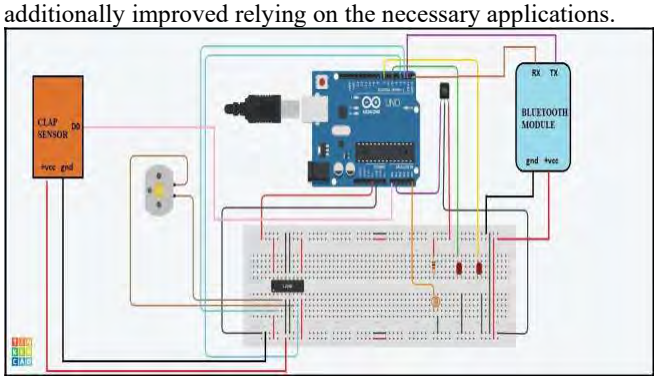


Fig.01 Circuit Diagram *A. Feasibility Study*

In this step, check whether the proposed frameworks are in fact possible or not. i.e., all the innovations required to build up the framework region practical promptly or not. Specialized Practicality decides if the association has the innovation and aptitudes important to do the task and how this ought to be acquired. The framework can be doable on account of the accompanying grounds:

- All essential innovation exists to build up the framework.

- This framework is excessively adaptable and it very well may be extended further.

- This framework can give assurances of exactness, convenience, dependability and the information security.

- This framework can give moment reaction to ask.

- Our venture is in fact doable in light of the fact that, all the innovation required for our undertaking is promptly accessible.

microcontroller. Arduino Uno has 14 advanced inforaation/yield pins (out of which 6 can be utilized as PWM yields), 6 simple information sticks, a USB association, A Force barrel jack, an ICSP header and a reset button.

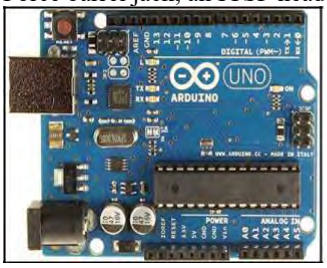
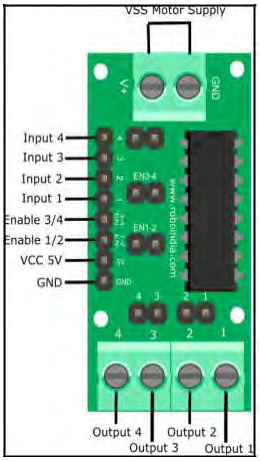


Fig.02 Arduino UNO Fig.03 IR Sensor

IR Sensor (Fig.03): As referenced that the model is IR sensor situated it deals with some standard. Infrared sensors deal with the guideline of reflected light waves. Infrared light reflected from articles or sent from an infrared remote or signal. Infrared sensors are additionally used to quantify separation or closeness. The reflected light is recognized and afterward a gauge of separation is determined among sensor and article.

Motor Driver (Fig.04): The Motor Driver is a module for engines that permits you to control the working velocity and course of two engines all the while. The Motor Driver which we are utilizing is structured and created dependent on L293D IC.



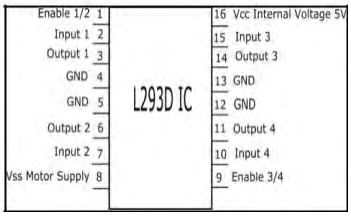


Fig.04 Motor Driver Fig.05 L239D IC

L239D IC (Fig.05): The L293D is a 16 pin IC, with eight pins, on each side, to controlling of two DC engine at the same time. There are 4 Info pins, 4 Yield pins and 2 Empower pin for each engine

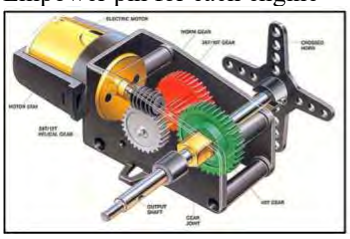


Fig.06 Servo Motor

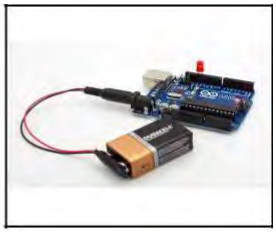


Fig.07 Battery

B. Implementation Details

Arduino Uno (Fig.02): It is a microcontroller board dependent on 8-piece ATmega328P microcontroller. Alongside ATmega328P, it comprises of different segments, for example, precious stone oscillator, sequential correspondence, voltage controller, and so forth to help the

Servo Motor (Fig.06): The servo motor position can be controlled more correctly than those of average DC engines, and for the most part, they have three wires like force, GND, and control. Capacity to these motors ceaselessly applied, with the servo engine control circuit changing the attract to drive the servo engine. These engines are intended for increasingly

careful assignments where a motor position should be clear absolutely like moving a mechanical arm or controlling the rudder on a pontoon or robot leg inside a specific range.

Battery (Fig.07): The Arduino Uno can be fueled by means of the Smaller than usual B USB association, 6-20V unregulated outer force supply (pin 30), or 5V directed outside force supply (pin 27). The force source is consequently chosen to the most elevated voltage source. So you can interface battery positive to VIN pin and negative to Ground pin legitimately.

C. Features of the project

The project contains the following features:

- To create a working model of an obstacle avoiding car, having the following features:
- Damage control
- Remote operated
- Reverse mechanism
- Obstacle avoidance
- Dead-end avoidance

D. Methodology used Algorithm Step 1: Start

Step 2: Initialize IR Sensor, Servo Motor, Motor Driver and Batteries

Step 3: If Obstacle detected on the path then

Move backward and produce a beep sound by piezo buzzer

Else Move in the same direction

Step 4: If we come across a Dead end on the path then

Move backward and produce a beep sound by piezo buzzer.

Else Move in the same direction Step 5: Stop

IV. RESULTS AND DISCUSSIONS

A. Output

Based on the sensor use, we are able to achieve the desirable outputs as follows:

1. As an when infrared sensor detects the obstacle it acts immediately and gives response to Arduino board

2. The piezoelectric sensors produces a beep sound in response to the obstacle activity

3. The infrared sensor also detects the dead end it uses a call back mechanism and the robot goes back as response.

B. Outcome

Phase 1 Analysis: Creating the synopsis, literature survey

and feasibility study for proposed solution which will act as an input to the design phase.

Phase 1 Planning: Proposing Statement of work, scope definition and scope boundary for definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done.

Phase 1 Designing: Proposing design architecture of Arduino and Control using IOT system for implementing the surveyed case studies and synopsis.

Phase 1 Implementation: Various modules will be created like connection of sensors. Completion of coding Collecting data from sensors

Phase 2 Testing:

Case 1: IR blaster remote connection Components/things required: IR blaster remote, Arduino UNO, IR sensors after connection as shown in the circuit diagram, the car was operated with the help of remote, and it was successfully moving as instructed. And whenever any obstacle encountered the car was automatically going into halt state. The IR blaster remote was tested successfully and was in working condition.

Case 2: IR sensor. Components required: IR sensor after connection as shown in the circuit diagram, the predicted results were successfully observed. Whenever the obstacle encountered or if deep distance between the car and the ground surface encountered the car successfully stopped.

Case 3: Sound Transducer producing alert signal: Components required: IR Sensor, Sound transducer after connection as shown in the circuit diagram, they predicted results were successfully observed. Whenever the obstacle encountered at that time the sound transducer was generating sound as an alert signal.

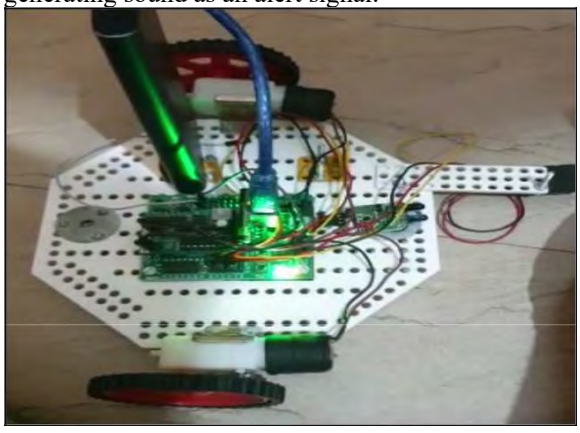


Fig.08 System Design

V. CONCLUSION

The code was transferred utilizing Arduino IDE and the code ran effectively. The circuit outline was made utilizing Tinker lowlife application and the circuit was effectively associated alluding to the circuit chart obstruction staying away from robot was effectively executedand was proficient to utilize. We manufacture an automated vehicle which moves in various ways like Forward, Backward, Left, and Right when information is given to it. The objective of our venture is to make an independent robot which cleverly recognizes the obstruction in his way and explore as indicated by the activities that we set for it.

References

[1]. D.S Vincent, Dr. S. Karthikeyan, E. Manokaran; Anticipation of mishaps in broad daylight transport enterprise at Madurai city Shraddha Dhonde..International Journal of Science, Engineering and Technology Research (IJSETR), Volume 5, Issue 2, February 2016 (http://ijsetr.org/wpcontent/uploads/2016/02/IJSETR-VOL-5-ISSUE-2-439-442.pdf).

[2]. Ranik Guidolini, Claudine Budae, Mareilla Berger; Mohammad Nasucha . 55 – jsisko JURNAL SISTE KOMPUTER – Vol. 5, No 2, November 2015, ISSN : 2087-4685, e-ISSN: 2252-3456(https://www.researchgate.net/publication/295519484_DEVELOPMENT_OF_AN_OBS TACLE_AVOIDING_ROBOT).

[3]. Rakesh Chandra Kumar , Md. Saddam Khan International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 4, April 2013 (https://www.ijareeie.com/upload/april/29_OBSTACLE.pdf) OSForensics: Accessed online: https://www.osforensics.com/index.html

[4]. Widodo Budihart Computational Intelligence and NeuroscienceVolume2015,ArticleID745823 (https://www.hindawi.com/journals/cin/2015/745823/)

[5]. Available online through the following URL: http://www.aurdino.cc/

[6]. Available online through the following URL: http://www.electronicsforu.com

Smart Gas And Flame Leakage monitoring System

Taher Yusuf Ali, Yash Yadav, Aakash Gupta
TCET,Mumbai
taherwasanwala@gmail.com, ygyadav1423@gmail.com, gupta29470@gmail.com

Abstract - **Internet of Things, is the currently the most trending field in the Branch of Information Communication Technology. In case of construction areas, residential and commercial buildings, there are high chances of domestic fire. Moreover, mining industry being the most fire prone area, there is requirement of a system that can generate information of the region to the miner on the basis of the gas getting generated while mining. Gas is generated first and then fire. The objective of this project is to build a gas and fire leakage detection sms alert system based on IoT technology. A Gas detector is a device which senses the gas and based on the Gas Quantity that is present in the atmosphere it will provide alerts to the beholder. A fire detector will sense the temperature above the threshold value and alert the owner by giving alert message. Nowadays, Gas Detection and sms alert System has been the basic requirement in various regions as the damage caused by the fire accidents is catastrophic. Also, as now the apartments are pretty close and congested, fire can spread very quickly. So, if gas detection and sms alert system is installed while building the apartment, the fire can be suppressed while it is in the small area and further loss of life and property due to fire can be prevented. The output of the gas detectors were been shown according to different test cases and thus the gas-flame detection and sms alert system become more responsive then those currently in use.**

KEYWORDS: Gas, Flame , Alaram,Threshold,SMS.

II. INTRODUCTION

Commercial buildings, Mining Regions, Factories, etc. Nowadays has become more prone to fire because of some or the other issues like short circuit, gas leakages etc. So there is a need of Gas and fire detection and sms alert system that can help us to overcome the damage to the lives and valuables. It is the fact that Gas is the first to come and then the fire. Hence detecting the fire at the gas level can help avoiding the fire to cause catastrophic destruction. The Gas detection and sms alert system is a Device that detects relevant amount of gas in the region and based on that it generates the output that there is a fire or not and based on the output of the further actions by fire- fighters and the administrations can be taken. Moreover now a days there has been an increase in this these days, this project also take an initiative to overcome that by discussing about it in the future scope.

III. BACKGROUND

Gas and Fire leakage detection has become a crucial aspect in design of buildings, both commercial and domestic, as opposed to about 70 years ago when automatic detection was rarely provided in buildings. Before introduction of gas and sms alert, fires resulted in the loss of human lives and damage of property and it was mainly attributed to lack of a mechanism for early detection of gas. Early developments in design of gas alarm began in 1922 with observations by D. Bhattacharjee [3] and later by Walter Jaeger in 1930. However, early gas detectors required high voltage power input. Further research was done and power requirement in gas detectors was reduced to make battery power viable and this made widespread installation in residences highly feasible. Later developments in gas detectors have sought to improve their performance, reduce power requirement and improve their nuisance alarm sensitivity and also to continuously monitor their status. The most recent advances in gas detectors have been motivated to make them “smarter”. Gas can be detected either optically (photoelectric) or by physical process (ionization). Detectors may use either or both methods.

IV. IMPORTANCE OF THE PROJECT

The requirement of the gas and fire leakage detection and sms alert system is because of the increase in the fire prone areas in the localized region where there is life like kitchen, buildings, construction region, mining industries petrol pumps etc. Hence, there is great importance of the project in day-to-day life. Importance of the Gas and fire Leakage Detection and Sms alert system are as below:

● Sms alert system are primarily designed to warn occupants
of a fire so they can safely evacuate the premises.

● Correctly maintained and operating sms alert system are effective and proven life saving devices.

● Failure to take advantage of this early warning, due to poor performance of gas detection system, has cost people their lives.

● Sms alert system are important in providing occupants of
buildings prompt warning if a fire occurs.

You need to carefully consider the benefits that you can achieve by reducing the number of unwanted alarm activations generated at your premises.

V. MOTIVATION

The various issues related to fire incidents occurred in the past year, i.e. 2017. Some of them are as:

1. The Saki-Naka fire incident that killed 12 workers, the fire incident at Omar Asif1[6] building where celebrities reside, and the Kamala Mills tragedy on a Friday night that killed 14 people – a blatant disregard for fire safety norms.

2. Be it slums, high-rises, restaurants, industrial units, studios or old buildings – each has witnessed massive fire incidents in this year. The violation of fire safety regulations has turned each into a tinder box, said experts.

Such incidents created a great loss of life and property both at the same time. And along with this due to the incident that happened in our classroom of AC caught fire. In this incident the air conditioner of our class first released some smoke in the environment and then it suddenly caught fire hence there was no such system that could inform about the fire. Hence, going through the above incidences and motivation to do something for the society we decided to develop a smoke alarm system using the IoT technology.

VI. OBJECTIVES

Tanvira Ismail said sometimes [4] it happens that fake leakage alarms are generated which can cause a loss of more than
$12,000 US dollars. So to avoid and tackle that we can use the camera and delay by 30 second in alarm which means if the leakage is detected to be above the threshold quantity, the detector will send the video to the operator and if the operator
approves that it is a legitimate fire in the region, the alarm will be horned and within 30 seconds of no reply from the user, the alarm will ring and accordingly the further actions can be taken.

• Leakage detection and alarm system can be used with quad copter and this quad copter can be used to detect the regions of the fire or going to catch the fire, thus can be used to protect things.

• Moreover it can be used by people in mining industries to help them for saving their lives.

It can also be used in F1 cars as there is high possibility of catching the fire when car runs at that speed there is high possibility of them catching fire due to the web based output it can be conveyed to the driver and thus he can save his life and win the race.

VII. PROBLEM DEFINITION

To design a microcontroller (IoT) based Gas & Fire Leakage detection and Alarm System that will continuously monitor the presence of significant amount of leakage and activate an alarm and based on that information about the region, solution can be provided.

VIII. PHASES OF PROJECT

● Phase 1: Planning, Analysis, Designing and Implementation. Analysis: Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation. Planning: After analysis we will first study about it and do some research on it for our better understanding of the project and also getting a rough picture about what would be our problem definition for the particular project. Designing: Then we will construct the design of the project and according to that, will list down all the requirements needed for the construction for the prototype of our project. Implementation: After acquiring the requirements we first developed the connections of the hardware components[7] (Arduino, MQ135 sensor,IR Flame Sensor etc.) and then added the coding part later with help of Arduino data cable to the Arduino.

● Phase 2: Testing and Deployment. Testing: After the prototype is ready we will first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding to it and will check again. Deployment: After, complete integration and testing of project real time running and operation of the system will be done. User need to check the value.

IX. FEATURES OF PROJECT

The Project consists of the following features:

1. Safety Signs: For showing the safeness of the project there is use of a buzzer .where buzzer will show the signs of leakage. If buzzer rang then there is a need to be alarm else there is whole safety.

2. High sensitivity and fast response: The smoke sensor used is highly sensitive to the amount of smoke generated each point of time in the region as soon as it senses the smoke above threshold it starts the alert system.

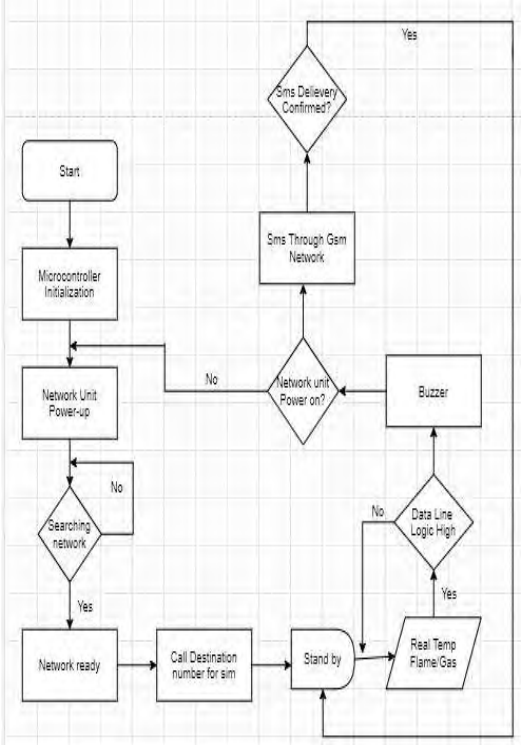
3. Long life and stable: The usual fire alarms are short lived because of the components used are of low grade here in this the sensor and the alert system does give an output to the Simple drive circuit.

4. The connections and the circuit design of this system is simple hence it is much more convenient to develop and find error in the system.

X. METHODOLOGY

The proposed method takes an automatic control action upon detection of gas. Initially if there is a Leakage detected then the electronic sensor i.e. the gas sensor that obeys the principle of sensor senses any fire in the region[5], if any leakage or fire is sensed then the output of this sensor goes high. This high signal is monitored by the microcontroller and it will identify the hike *Multicon-W 2020* in

the leakage amount. If there is a leakage or smoke detected, the consumer is informed about it through the LED light and



Sound of the buzzer.

Figure1: Flowchart of Proposed System.

XI. CIRCUIT DIAGRAM

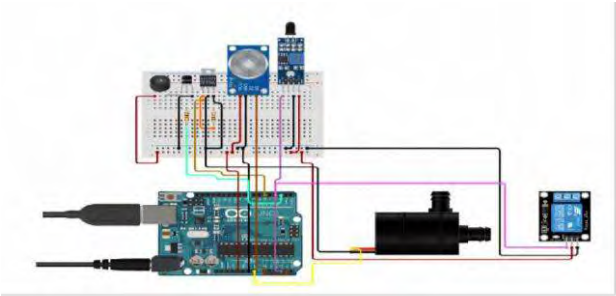


Figure2: Circuit diagram of Proposed System.

XII. RESULT & DISCUSSION

In this experiment, Flame and LPG gas are used for flame and gas leakage detection. Flame is detected by IR flame sensor and LPG gas is detected by MQ-135 sensor in various modes of parts per million**.** When the gas level was below the

threshold value the environment is safe and lcd shows the respective status.

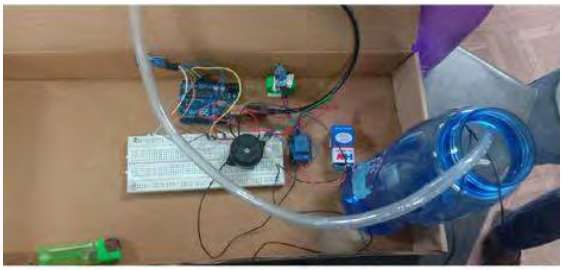


Image 1: Final System Orientation.

XIII. CONCLUSION

Through this system it's successfully shown the detection of Gas Leakage by IOT based project is possible which can detect Gas leakage and warn owners and remote contacts of an alarm situations. The alarms are detect reliably even in situations where there is high ambient noise. A weakness of the system is in the processing overhead used to constantly monitor the audio input of the phone. This application demonstrates the mobile phones ability to be used as a advanced warning device in the event of a smoke alarm activation, its automated group alert system reduces the chance of harm to people who have not been able to respond to an alarm.

XIV. FUTURE SCOPE

1. Gas Leakage detection and alarm system can be used with quadcopter and this quadcopter can be used to detect the regions of the fire or going to catch the fire, thus can be used to protect things.2. Moreover it can be used by people in mining industries to help them for saving their lives.

3. It can also be used in F1 cars as there is high possibility of catching the fire when car runs at that speed there is high possibility of them catching fire due to the web based output it can be conveyed to the driver and thus he can save his life and win the race.

References

[1] "Ashish Shrivastava, Ratnesh Prabhaker, Rajeev Kumar and Rahul Verma "GSM BASED GAS LEAKAGE DETECTION SYSTEM" International Journal of Technical Research & Applications, e-ISSN:2320-8163, Volume 1, Issue 2 (may- june,2013), PP. 42-45.

[2] D. bhattacharjee,P. Bhatnagar, S. choudhury, "Design and Development of a Flexible Reliable Smart Gas Detection System", IJCA,31(2011) 1-8.

[3] Tanvira Ismail, Devoleena Das Jyotiromoy Saikia, Jyotirmoy Deka "International Journal of Electronics and Communication Engineering & Technology (IJECET), ISSN 0976 – 6472Volume 5, Issue 5, May (2014), pp. 122-135.

[4] Omar Asif1*, Md. Belayat Hossain2, Mamun Hasan2, Mir Toufikur Rahman3,Muhammad E. H. Chowdhury4 "Fire-Detectors Review and Design of an Automated, Quick Responsive Fire-Alarm System Based on SMS" Int. J. Communications, Network and System Sciences, 2014, 7, 386-395.

IoT Enabled Smart Greenhouse using Arduino and NodeMCU

Jigar Vaishnav, Harshit Tripathi, Krishna Vishwakarma
Department of I.T.
TCET, Mumbai
jigar207844@gmail.com, tharshit07@gmail.com, mrkvishwakarma@gmail.com

Abstract—**Internet of Things (IoT), is currently the most trending field in the Branch of Information and Communication Technology. In the case of farming or agriculture, due to extreme weather conditions, the crops may get damaged and as a result, there are huge losses. Moreover, due to climate change the occurrences of floods, droughts, etc. which indicates that farming through traditional means will only increase loss in that field. A Greenhouse is a structure in which plants are grown in a closed controlled environment. Greenhouse warm up during the day when sun-rays penetrate through it, which heats the plant, soil and the structure. Greenhouses help protect crops from several diseases, particularly those that are soil-borne. Greenhouse temperature should not go below a certain degree, High Humidity can result to crop transpiration, condensation of water vapor on greenhouse surfaces, and water evaporation from the soil. To overcome such challenges, this system comes into rescue. After studying the literature survey, it was decided to overcome the research gaps and the system was tested using the test cases and then the outputs were generated. The output has been shown according to different test cases and thus the automated greenhouse system proved to be more responsive than those currently in use.**

Keywords – Internet of Things; Greenhouse; Sensors; Automation; Information and Communication Technology; Climate Change; Arduino; Programming;

I. INTRODUCTION

Internet of Things refers to the interlinking of ordinary everyday objects which are equipped with little or no intelligence, with software, electronics, network, and sensors which enables them to collect and exchange data. Internet of Things i.e. IoT enables us to automate certain processes and tasks which earlier required human intervention. Now a day's in our surrounding everything can be controlled and operated automatically. But there are some sectors where automation has not been adopted[1]. Agriculture is one such field. IoT will increase the ubiquity of the Internet by integrating every object for interaction via embedded systems, which leads to a highly distributed network of devices communicating with human beings as well as other devices[2].

A Greenhouse is an important part of agriculture. Mainly greenhouse is used to grow plants under controlled climatic conditions. Greenhouses which control and monitor many parameters like temperature, humidity, soil moisture and light intensity are important for better plant growth. In recent years, the greenhouse industry has got greater progress and improved agricultural labor productivity[3]. For controlling all these parameters and growth of plants without human involvement we can design automated greenhouse. Automation can replace human resources completely to yield quality products.

II. BACKGROUND

India demands at enhancing the adaptive capacity of the farmers so that they can cope with climate variability in the vulnerable districts to achieve climate-resilience and make it socio-ecological agriculture. Under NICRA, this technology is demonstrated and implemented in more than 151 climatically vulnerable districts of the country.

The rise in global temperature is not only causing drastic changes in climate but also it's contributing to the irregular rainfall patterns. Uneven rainfall patterns led to increased temperature, elevated CO2 content in the atmosphere are one of the important climatic parameters which adversely affects the crop production in India. Studies indicate that these weathering parameters influence (67%) to the growth compared to other factors like soil and nutrition (33%) during the cropping season. For every 1°F increase in the temperature, the crop yield falls by 3% to 5% in India. Hence, there is a need to study the dependence of the temperature on crop productivity, stability, yield, and quality to uplift the country's economy.

III. OBJECTIVES AND SCOPE

The objective is to implement a low cost, reliable and scalable Smart Greenhouse which will also be automated and it will also provide the various parameters of a plant online which can be accessed from anywhere which will help the farmers to a greater extent.

- To make a System which is self-sufficient and needs a minimum amount of human interference.
- To monitor the various critical parameters of the plant's environment.
- To grow any plant in an artificially controlled environment.
- To perform analysis on the data which is generated from the sensors and sent to cloud for improving the system's efficiency.

The scope of this system is to implement a Smart Greenhouse which can sense various environmental parameters like temperature, humidity, soil moisture,

water level, light intensity and upload these values to the cloud where it can be accessed via a smartphone.

IV. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware and Sensors Used:

- Arduino UNO
- NodeMCU
- Relay Module
- DHT11 (Temperature Sensor)
- Soil Moisture Sensor
- LDR
- Ultrasonic Sensor
- LED Light

A short description of some important hardware components is as follows:

Arduino UNO:

It is a microcontroller board based on 8-bit ATmega328P microcontroller. It consists of components such as crystal oscillator, serial communication, voltage regulator, etc. to support the microcontroller.

Fig.1 Arduino UNO

NODEMCU:

NodeMCU is an open-source LUA based firmware developed for ESP8266 Wi-Fi chip. By integrating functionality with ESP8266 chip, NodeMCU firmware comes with ESP8266 Development board/kit.

Fig.2 NodeMCU

Software Used:

- Windows 10
- ArduinoIDE
- Blynk

V. DESIGN AND FLOW

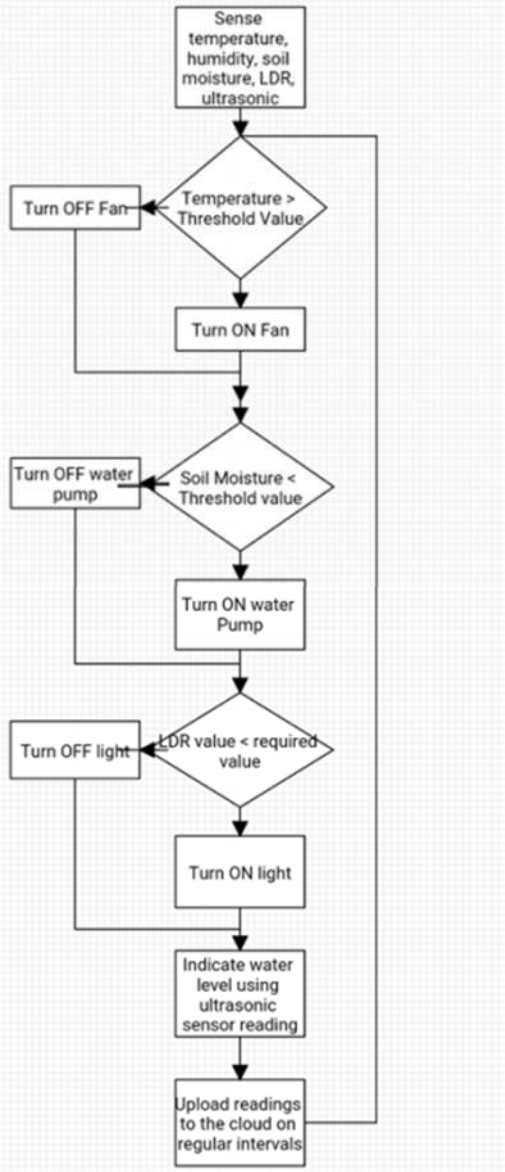


Fig.3 The Flow of System

VI. PLANNING AND METHODOLOGY USED

Phase 1: Designing

Proposing design architecture of Arduino and Control using IoT system for implementing the surveyed case studies and synopsis. Various modules will be created like the connection of sensors and actuators. Completion of coding and collecting data from sensors.

Phase 2: Implementation

i. DHT11 sensor is used to sense Temperature & Humidity inside the greenhouse. If the temperature is above the required value, the fan is turned on.

ii. The soil moisture sensor senses the moisture content of soil & turns on the water pump if the moisture is low.
iii. The LDR is programmed to detect shadow. This reading is used to turn on LED grow light if it is too dark.
iv. The Water Level of the tank is indicated using an Ultrasonic Sensor.
v. All these readings are sent to the cloud using a Wi-Fi connection by NodeMCU

Phase 3: Testing

i. In the first phase of testing, the readings of the sensors are obtained to check if there are any ambiguous readings. The soil moisture sensor is calibrated by taking a value in dry soil and another in wet soil. The sensors were tested successfully and were in working condition.
ii. In the second phase, the actuators are connected with the relay module which is in turn connected with the Arduino. Relay completes the circuit when a signal is sent by the Arduino. The threshold values are set to automatically trigger the actuators when the readings go beyond certain values.
iii. In the third and final phase, NodeMCU is connected to the Wi-Fi and the sensor readings are uploaded to the cloud at regular intervals.

VII. RESULTS AND DISCUSSIONS

Output: Based on the sensor use, we can achieve the desired outputs as follows:

1. As the DHT11's temperature reading goes beyond the threshold value, Arduino sends a signal to the relay module which turns on the fan of the Greenhouse.
2. When the soil moisture sensor value goes below the moisture limit which is required by the plant, the water pump is switched on by the relay module.
3. Also when the analog reading of the Light Dependent Resistor goes below the limit (which indicates shadow), the grow light of the greenhouse is switched on.
4. The readings of every sensor are uploaded to the cloud by the NodeMCU at regular intervals and a real time visualization is generated which can be accessed from anywhere in the world using a smartphone.

A screenshot of the dashboard inside the Blynk app interface is given below. Here we can keep track of the current conditions inside the greenhouse.

Fig.4 Blynk App Interface

Along with the functionality to keep track of the conditions inside the greenhouse, a graphical interface was also provided which displayed the variation of different weather parameters with respect to the chosen time duration.

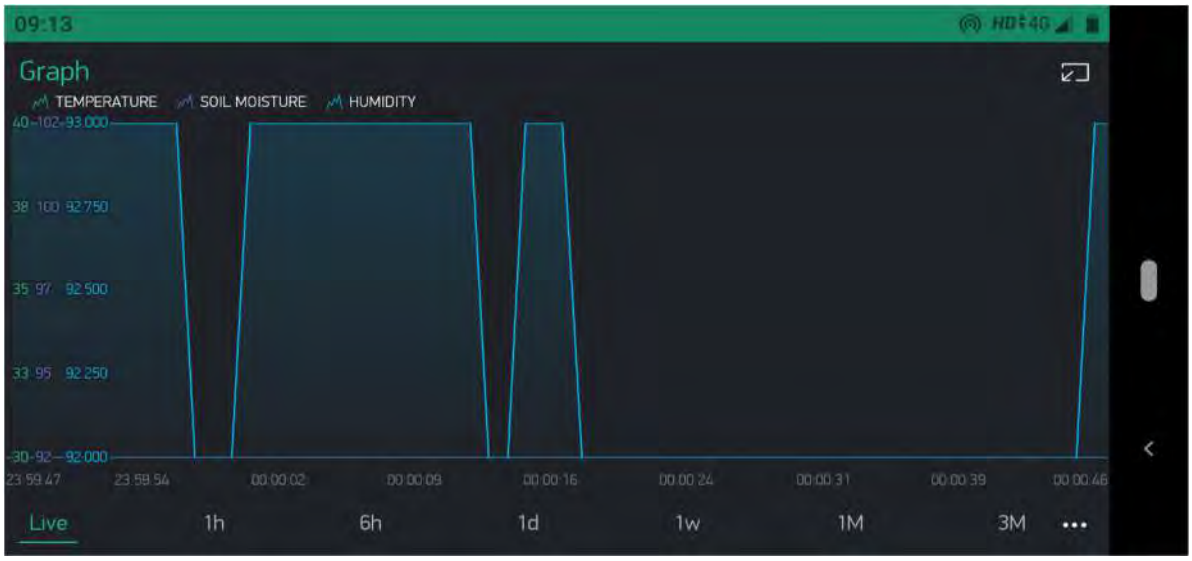


Fig.5 Graphical Representation

The proposed system is designed to function independent of human intervention. It can be programmed to maintain different weather conditions irrespective of the natural conditions. This enables us to grow a wide variety of crops which require a specific environment. Any anomaly in the system can be detected by just glancing at the dashboard.

Smart greenhouses have a very vast field of application. They enable farmers to grow crops which require certain specific climatic conditions. The farmers previously were majorly dependent on rains and other natural factors. These factors vary significantly every successive year which gives an uncertainty in the yield of crops. Greenhouses can help overcome these factors by simulating specific climate inside them. This enables farmers to grow crops which are not native to their region. Smart greenhouses will help collect data which when analyzed can give us some very valuable insights on the changing environmental factors inside the greenhouse and can help us build better greenhouses in the future. Smart Greenhouses are a step towards a smarter future.

VIII. CONCLUSION

In this project, we created a working model of a Smart Greenhouse by carefully analyzing all the parameters in an automated greenhouse. This model performs several functions which are as follows:

- Water is supplied to the plant when the water content of the soil is low.
- The fan is switched on to bring the temperature down as soon as it goes beyond a limit.
- When the amount of light available to the plant decreases, the LED light is switched on.
- The water level of the tank is also indicated.
- Sensor data is regularly uploaded to the cloud.

The above points indicate that our model is working as intended. The greenhouse is performing all the functions as it was planned earlier. This project will act as a contribution to the field of smart farming. Hence, we have successfully achieved our objective to create an automated greenhouse which is IoT enabled.

Importance of Project:

1. Use new technology that uses the Internet of Things (IoT) concept.
2. To grow any plant in an artificially controlled environment.
3. To achieve cost savings.
4. To enhance the ability of a device to research by collecting data.
5. To automate the process of farming.

References

[1] Payal V. Jadhav, Sonali G. More, Ravina S. Shinde and Rakhi S. Pagar, “IoT based Automation in Green House,” International Journal of Research in Engineering, Science and Management Volume-2, Issue-2, February-2019.

[2] Feng Xia, Laurence T. Yang, Lizhe Wang and Alexey Vinel, “Internet of Things”, INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS Int. J. Commun. Syst. 2012; 25:1101–1102.

[3] Yongxian Song, Juanli Ma, Xianjin Zhang and Yuan Feng , “Design of Wireless Sensor Network-Based Greenhouse Environment Monitoring and Automatic Control System” JOURNAL OF NETWORKS, VOL. 7, NO. 5, MAY 2012.

[4] Edin Mujčić and Una Drakulić, “Monitoring and Control of the Greenhouse Using Remote Control Device and IoT”, International Journal of New Innovations in Engineering and Technology Volume 12 Issue 3 December 2019.

[5] Vidhya Shree A , Supriya R and Gayathri J, “GREENHOUSE MONITORING USING ARDUINO PLATFORM”, IJARIIE-ISSN(O)-2395-4396 Vol-5 Issue-2 2019.

Rfid Access Control System

Omkar Punekar, Prasad Paradkar
Tcet, Mumbai
kartikpunekar7@gmail.com, prasadparad37@gmail.com

Abstract – **There has been raising interest for secure system that must be reliable and fast react for the businesses and friends. RFID (Radio Recurrence Identification) is one of the steady and quick methods for recognize the material article. In the sometime in the past the scanner tags are progressively ideal when contrasted with RFID in light of their expense however now daily's RFID are effectively accessible and are increasingly advantageous to utilize. Research has rolled out some extreme improvements which makes its programming significantly shorter and simpler is a direct result of supplanting microcontroller with Arduino. Arduino makes the circuit and programming significantly more obvious. Paper depends on security get to utilizing RFID and Arduino module**.

Index Terms— RFID tags, Arduino UNO, RFID Reader.

I. INTRODUCTION

This paper gives information on radio recurrence recognizable proof (RFID) innovation. At first RFID labels were made to in the long run supplant scanner tags in various chains. Their points of interest are that they can be perused remotely and with no observable pathway, contain a bigger number of information than scanner tags, and are more grounded. As the paper depicts The on-going innovation, incorporate the recurrence ranges utilized and gauges required. With the expansion in universality

of RFID labels, in any case, protection became unease. The paper diagrams plausible assault that can conflict with one\'s security and it additionally portrays negate measures. The RFID innovation didn't stop at thing-level labelling. Since the utilizations for RFID labels are so broad, there is an enormous enthusiasm for bringing down the expenses for generation of RFID labels. Things being what they are, printing labels may turn into a conceivable option in contrast to customary creation.

II. BACKGROUND

Today individuals are confronting extra issues approximately security in all finished world, nowadays security is the maximum fundamental difficulty anywhere on the planet; so, safety of the whole thing increases increasingly elevated importance as of overdue. Here in this paper, endeavoring to copy the thorough writing study recognized with the one-of-a-kind entryway bolts and gate protection frameworks that are critical within the fields, for example, home, agencies and vehicle safety where potential outcomes of invasion are increasing step by means of step. In beyond days, the exploration is long gone onspecific entryway bolt safety frameworks like normal safety frameworks which give signs and symptoms utilizing caution. Because of the headway in past due strategies, some entryway bolt protection frameworks rely on microcontroller, GPS, numerous sensors, programming like MATLAB, PROTEUS, and biometrics like face acknowledgment, Iris scanner, RFID, Smart Card and mystery key and so on. Every framework has its own factors of hobby and weaknesses. In the greater a part of frameworks, SMS method is utilized for correspondence so the framework will move in the direction of turning into savvy, extra dependable and it'll require less investment to bring message. As security ends up real trouble in recent times, the security checking frameworks today desires to make utilization of the latest innovation. In some papers, the creators have added entryway bolt protection looking at framework in mild of inserted and Zigbee and occasionally the bolt is ensured via programmed secret key henceforth it couldn't without a good deal of a stretch hack through programmers. Additionally, the upgraded safety frameworks are on hand in mild of android stage, remote methods and inserted frameworks. A amazing deal of adjustment takes places in special Gate bolt security from the maximum recent couple of years, in subsequent coming years several progressions will happens. RFID, Radio Frequency Identification is a modest innovation, may be carried out for some applications, for example, protection. RFID (Radio Frequency Identification) is one of the predictable and brief methods for distinguish the material protest. In the long-earlier the standardized tags are more perfect while contrasted with RFID in light of their cost however now daily's RFID are effortlessly accessible and are more helpful to utilize. Research has rolled out a few brilliant upgrades which makes its programming a considerable.

III. OBJECTIVES

The objectives of project:

It will also use to monitor the access to door.

It can reduce chances of entry of trace passer (unknown person).

It can avoid time consumption of manual attendance records

It improves whole system by making it online and high-tech. It helps to find who and when accessed the system.

IV. PROBLEM DEFINITION

Security represents protection of our lifestyles and assets. Ensuring protection of peoples and their precious things is very vital for the prevention of illegal handling. Hence, especially specializing in door lock protection or gate security is very crucial to keep away from the further troubles in monitored area. Even with the usage of mechanical locks, the crime, robberies get happened because of the reality that such locks were easily broken. So, there's a want to invent other type of locks which cannot be easily broken. So, we through our venture we are presenting a lock that is a RFID Door Lock that could be a easy door lock that can be connected to any ordinary door. It unlocks when the presence of the RFID tag is close by and locks when it is not detected. There is also a LED that lets the user recognize while the door is locked and

unlocked. It provides the safety of a normal door lock without the problem of keys.

V. SYSTEM COMPONENTS

RFID tags: A RFID tag is a smooth card of Visa size which is perused by using a RFID according to person. It works at 125kHz and accompanies a one in every of a kind 32-piece ID. Ordinarily, every tag has top notch ID range which cannot be changed. We can discover its thrilling ID thru programming.

Arduino Uno Board: It is an open supply hardware prototyping stage dependent on bendable, simple to-utilize gadget and programming. It is proposed for craftsmen, fashioner, experts and anyone eager on producing specific shape for objects or ecological reason. Arduino UNO is a board primarily based on ATmega328 microcontroller. It contains of 14 computerized input/yield pins, six easy statistics sources, a USB connect for programming the on-board microcontroller, manage jack, an ICSP header and a reset catch. It is paintings due to 16MHz precious stone oscillator and consists of everything predicted to help the microcontroller.

RFID Scanner: A RFID consistent with-consumer is the cerebrum of the RFID framework and is vital for any framework to work. Per -users, likewise called questioners, are gadgets that transmit and get radio waves so as to speak with RFID labels. RFID in step with -customers are often partitioned into unmistakable sorts – Fixed RFID Readers and Mobile RFID Readers. Fixed per users stay in one location and are regularly set up on dividers, on work areas,

into entryways, or other desk bound areas. A regular subset of constant in keeping with customers is coordinated per users. A coordinated RFID according to user is a consistent with consumer with a worked in receiving twine that usually incorporates one greater reception equipment port for the association of a discretionary out of doors radio twine too.

Buzzer: We are utilising piezoelectric bell. This is an admonition/signal that invalid endeavour is achieved to get right of entry to framework.

Servo motor: It can be utilized to expose show off of entryway or door opening.

LED: The LED motive is whilst a correct entry is made it glows.

Arduino IDE: The IDE is a special program strolling on a laptop that lets in one to write down sketches for the Arduino board in easy language version after processing the language.

VI. METHODOLOGY

The undertaking has the accompanying work process: on landing at the entryway where the entrance control is introduced, one is approached to surmised their RFID tag to the per user. The per user peruses the tag and the microcontroller think about the tag's UID for match and award get to if there is a match and deny get to if there is no match. A RFID tag can be included or expelled through the Arduino IDE or some other programming language that Arduino gets it. For changes made on the sketch (for example including or evacuating a tag) to be viable on our framework, the sketch must be re-transferred to the Arduino board to

supersede past sketch. In other to include another client, we should initially enrol the client with the framework at that point, relating client data is consume in RFID tag. The new label will presently be open through the framework.

VII. BLOCK DIAGRAM

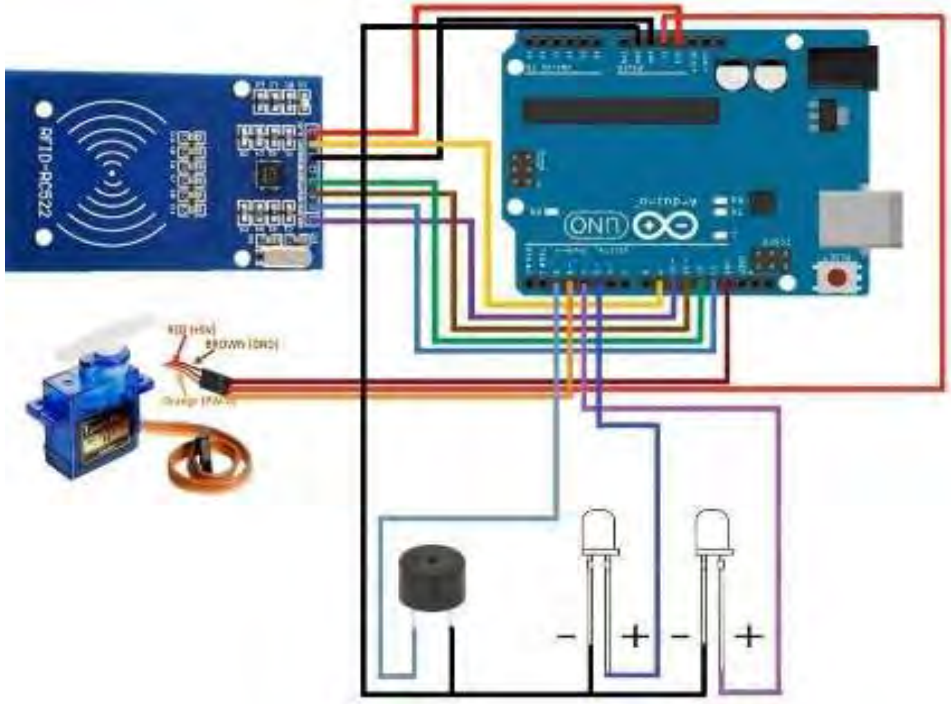
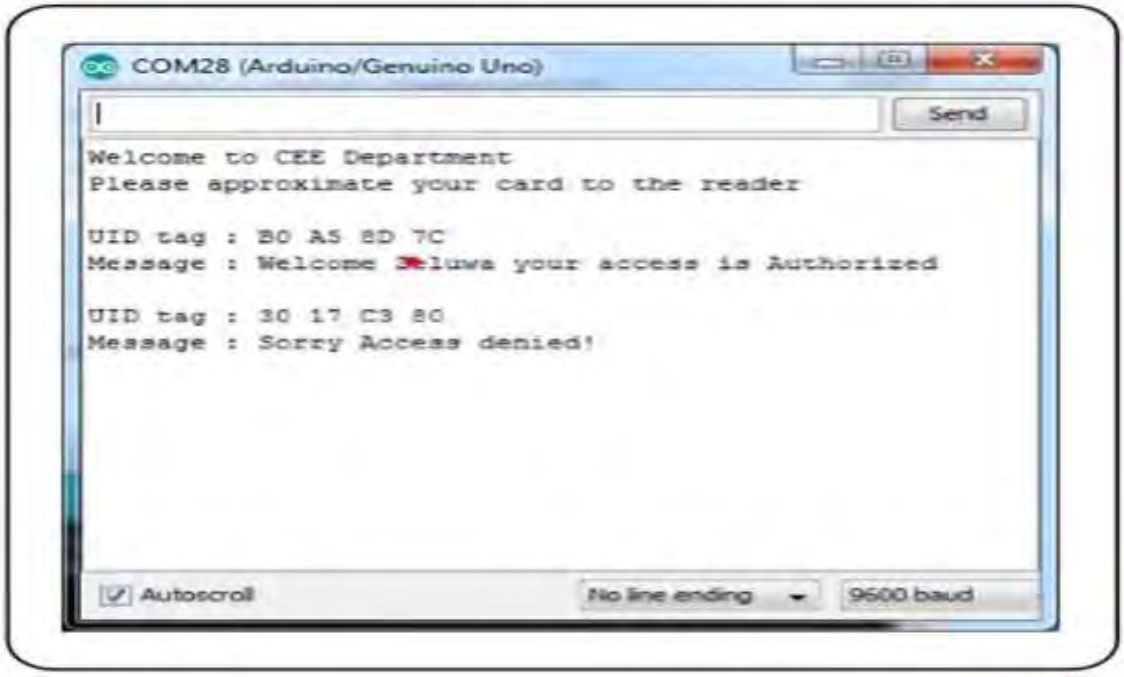


Fig1: Block Diagram

VIII. GUI SCREENSHOT



IX.

Fig 2: GUI Screenshot

X. OUTCOMES

In this project, a prototype of automatic access control system for use in an environment is presented. The system uses radio frequency identification (RFID) with Arduino technology to differentiate between authorized and unauthorized users. The RFID reader reads RFID tag issued to the user and matches it with stored UID on the Microcontroller. On a successful match, the microcontroller grants access or deny access if no match was found. An automatic access control system using

Arduino and RFID has been prototyped and functioned as desired. The system can be installed at the entrance of a secured environment to prevent an unauthorized individual access to the environment. The Arduino board became programmed through Arduino IDE. All

necessary information about all customers is stored within the gadget.

XI. TESTING & DEPLOYMENT

The Arduino board became programmed through Arduino IDE. All necessary information about all customers is stored within the gadget. In other to add a new user, we should first register the consumer with the machine then, corresponding consumer statistics is burn in RFID tag. The new tag will now be accessible via the device. Access control device changed into analyzed using the following criteria: cost, electricity consumption, speed, user satisfaction, and stability. The improvements come within the location of cost, strength consumption, speed, and stability.

XII. FUTURE SCOPE

An RFID access control management system turned into designed for organizations. Ideas had been supplied for making whole system online the usage of portable tool and 3G cellular technologies. The future expectations from this task are to in reality implement such device for one or more agencies if sufficient budget is supplied to us.

XIII. CONCLUSION

The RFID enterprise goes to go into an energizing period wherein expanded selection will provide the manner to innovation providers to put resources into new, energizing developments. Alongside the brand-new improvements depicted above, headways in materials, herbal polymers, Nano innovation, and extraordinary territories will trade the manner RFID is fused into items. It relies on how unique one should be to enhance the utilization of this venture.

References

1. Anil K. Jain, Arun Ross and Salil Prabhakar. An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image and Video Based Biometrics, Vol. 14(1), January,2004.
2. Technology Intelligent Home: SMS Based Home Security System with Immediate Feedback International
3. Journal of Advance Research In Science And Engineering http://www.ijarse.com IJARSE, Vol. No.2, Issue No.5, May, 2013 ISSN-2319-8354(E).
4. "Proteus PCB Design & Simulation software – Lab center Electronics", Labcenter.com, 2017. [Online]. Available: https://www.labcenter.com/.
5. R. D. H. Arifin “Door automation system based on speech command and pin using android smartphone,” in 2018 International Conference on Information and Communications Technology (ICOIACT), March 2018, pp. 667–672.
6. Peng-Loon “NFC Smartphone Based Access Control System Using Information Hiding,” (ICOS), December, 2013

Wireless Notice Board

Apurva Dalvi, Shweta Mishra, Fatema Ujjainwala
TCET,MUMBAI
apurvapdalvi@gmail.com, shwetamishra5499@gmail.com, fatemaujjainwala97@gmail.com

Abstract— **The Internet of things (IoT) is currently the most trending field in the Branch of Information Communication Technology. It is the extension of Internet connectivity into physical devices and everyday objects. IoT is a rapidly expanding field in the world of technology. It helps ease life for a person by means of internetworking physical devices, automotive, home appliances and other items embedded with electronics, software, sensors and actuators.**
Nowadays conveying messages on a large-scale using notice boards is used widely from schools to various organizations. The significance of notice boards in public areas like bus stands, railway stations, airports, and banks, etc. is known by us. But changing these boards every day is a laborious task and a waste of time. At present, all electronic boards are designed with a wired system. The major drawback of designing these boards is; not flexible and cannot be located anywhere due to messy wire.

To overcome this problem, a wireless board is designed to display the latest information. The main objective of this project is to develop a solution that provides real- time notices; so, a Wireless Notice Board using Arduino and GSM technology was developed that displays messages sent from the users mobile. It has a very wide scope rather than just being a simple notice board.
Keywords – Arduino, GSM, Wireless Notice Board, Messages.

I. INTRODUCTION:

The Internet of things (IoT) is the extension of Internet connectivity into physical devices and everyday objects. IoT is a rapidly expanding field in the world of technology. It helps ease life for a person by means of internetworking physical devices, automotive, home appliances and other items embedded with electronics, software, sensors and actuators.
All present electronic notice boards are designed with a wired system. The major drawback of these boards is that, it's not flexible and cannot be located
anywhere due to messy wire. Also, some of the places need urgent notices like in college, railway stations share-market, and this notice should be in real-time; so we need a real-time notice. So to provide a solution that displays notices in real-time, a wireless board is designed to display the latest information using the GSM technology.

GSM network is widely used today whether it is for calling or SMS. When a mobile user sends the information from his mobile, it is received by a SIM slot, which is integrated into a GSM modem at the receiver end. Once the information is received, it is displayed on the notice board. It has a very wide scope rather than just being a simple notice board.

II. BACKGROUND:

Traditional Notice Boards are often ignored as there is a lot of information on the board. They can become untidy and irrelevant as time passes and it is difficult to maintain and regularly update the information on the board. Also, when the information on the notice board is to be displayed, a hardcopy is required that leads to the usage resources like paper and finances that are recursive.

The alternative for Traditional Notice Boards is a Hard-Wired Digital Notice Board. These Notice Boards decrease the recurring financial expenditure. But whenever the information on this notice boards is updated it has to be done from an allotted system you cannot update the information from remote locations or any other location except the allotted one.

As there is a need of real-time update of information, a wireless board is designed to display the latest information using the GSM technology. This allows us to display notice in real-time by sending the information from a mobile, which is received by a SIM slot that is integrated into a GSM module at the receiver end. Once the information is received, it is displayed on the notice board.

III. IMPORTANCE OF THE PROJECT:

The requirement of Wireless Notice Board is because of the need of real-time notices in some of the places like in college, railway stations share- market, etc. which require urgent notices. Hence, there is great importance of the project in day-to-day life. Importance of the Wireless Notice Board are as below:

The Notice that is displayed on the LCD screens is repeated unless there is a new notice that is to be displayed.

Correctly maintained and constantly updated notices are effective and proven to benefit the Organization.

Constantly updated notices keep the people updated about the situations.

IV. OBJECTIVES AND SCOPE OF THE PROJECT:

The main objective is to design an automatic self-enabled highly reliable electronic notice board, a display connected to a server system should continuously listen for the incoming call from client or user process it and display it on the LCD screen message displayed should be updated every time the user sends new data only

authenticated people should be able to able to access the server.

The proposed system provides efficient ways to outspread notices through LCD Screen. The LCD Screen is interfaced with GSM module through Arduino.

When we send SMS from mobile phone to GSM module then GSM receives that SMS and sends it to Arduino. Now Arduino read this SMS and extract main notice message from the received string and stores in another string.

And then sends the extracted message to 16x2" LCD by using appropriate commands. Which is displayed on the LCD Screen.

V. SUMMARY:

Due to advances in wireless communication it is possible to control devices without physical movement. This is possible with the use/design of an embedded system in communication giving comfort and safety to human life. The main aim of our project is to design a Wireless Notice Board that can replace the currently used programmable electronic display. The Receiver and display board can be programmed from an authorized transmitter. The microcontroller receives the SMS and displays the desired information. Such a system is helpful for immediate information transfer

VI. PROBLEM DEFINITION:

To design a Wireless Notice Board using Arduino and GSM Technology that will display urgent notices, information in real-time that are sent by user's mobile, which can replace currently used conventional notice boards.

We will use various components such as GSM 900a Module and the 16*2 LCD for displaying the received SMS, Arduino UNO as a processor which turns out to be the best and most suitable option, Breadboard and Jumper Wires to make connections among all the components. Overall the system provides a complete solution for the problem.

VII. RESULTS:

A. Outputs:
When all the components are connected properly, the LCD shows the respective status.

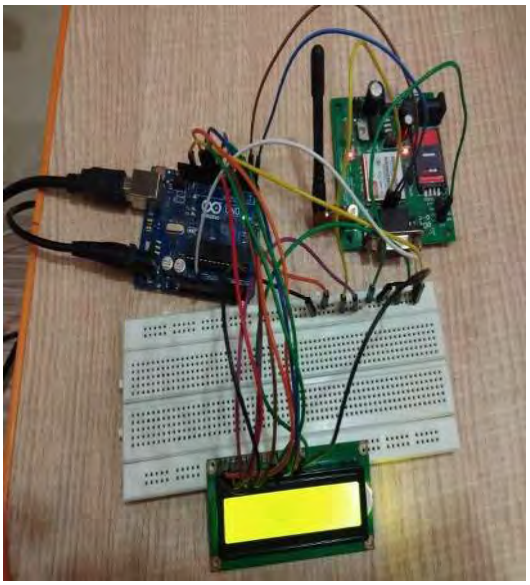


Fig1: All components connected together

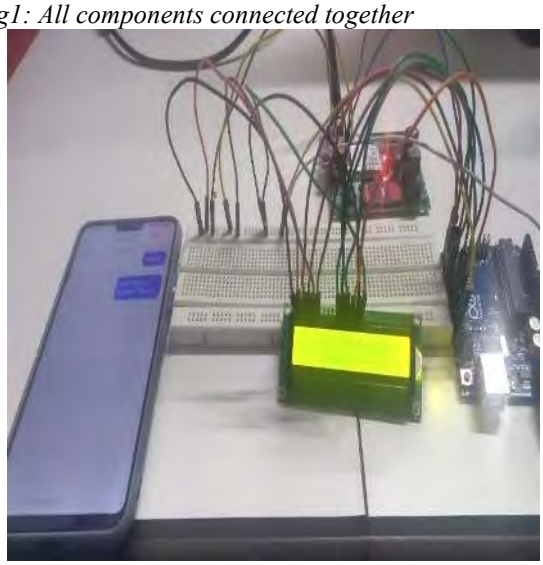


Fig2: Working model of the project

VIII. OUTCOMES:

Phase 1:
1.Analysis:
Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase.

2.Planning:

Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done.

3.Designing:

Proposing the design architecture for the prototype.
4.Implementation:

Implementation of Arduino UNO with the help of code running in the Arduino IDE environment.

PHASE 2:

1.Testing:

IX. SCREENSHOTS:

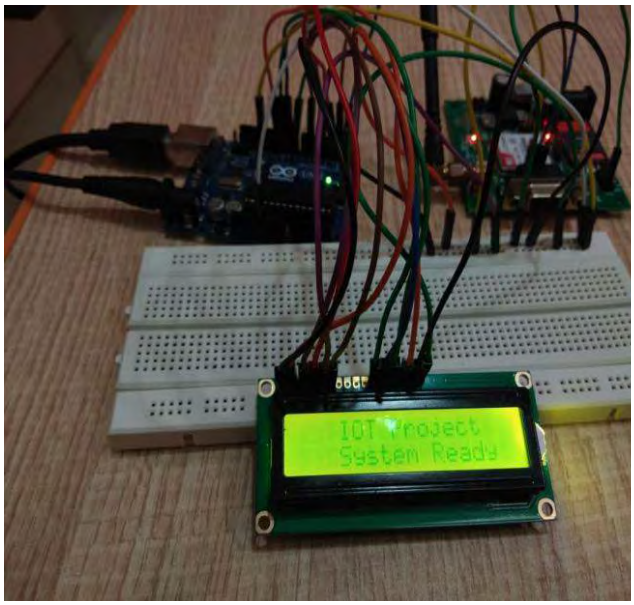


Fig 3 : Wireless Notice Board

Fig 4 : SMS interface of the project

Testing was done by sending messages multiple times to the GSM module.The testing of the model successes are mentioned in the above figures.

X. IMPROVEMENTS IN RESEARCH GAPS (FINDINGS):

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. They are as follows:

We can integrate Artificial Intelligence with the existing system, which will analyze the data and situations of the surroundings and display urgent notices on its own. For instance, Displaying the urgent message “Evacuate Building” in case of fire emergency.

We can use Wi-Fi Module instead of GSM Module, to increase the efficiency of the system.

We can use LED Screens to display notices, which might reduce the power usage.

XI. FUTURE SCOPE:

Electronic Notice Board is one of the applications where WIFI and Raspberry pi can be used effectively. It can also be used in Malls and Highways for Advertisement purpose. A moving display with variable speed can also be used in place of static display.

1.In this project we are sending messages over a GSM network and displaying it on an LCD by the use of AT (ATtention) commands.

2.The same technique can be applied to control electrical appliances at distant locations.

3.Robots can be controlled in a similar manner by sending the commands to the robots. This technique can be used for spy bots at distant locations, utilized by the defense organizations to monitor the movement of enemy troops.

4.Alphanumeric LCDs have limitations on size as well as number of characters. These can be replaced with large LED displays which are not only eye catching but display characters in a moving fashion one after the other.

XII. CONCLUSION:

As the technology is advancing every day the display board systems are moving from Normal hand-writing display to digital display. Further to Wireless display units. This paper develops a wireless notice board system with GSM modem connected to it, which displays the desired message of the user through an SMS in a most populated or crowded places like railways, bus stations, airports and also at roadside for controlling the traffic and in emergency situations. It is economical system and very easy to handle. Use of papers in displaying of notices is avoided. The information can only be updated by the authorized persons.

XIII. ACKNOWLEDGMENT:

The success and outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have completed the project successfully. I would like to thank everyone for their guidance.

We sincerely thank our principal, Dr. B. K. Mishra, Vice Principal, Dr. Deven Shah, Mentor Dean, Dr. Kamal Shah and HOD, Dr. Rajesh Bansode for always encouraging us to do our best. We highly indebted to our

guide Mr. Anil Vasoya, who supported and constantly supervised us throughout the project and helped us in not only completing this project but also provided us with ample amount of knowledge that was beneficial to us.

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of IT Department who helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in laboratory for their timely support.

We would like to express our gratitude towards our parents for their kind co- operation and encouragement which helped us in completion of this project.

References:

[1]. Guifen Gu and Guili Peng The Survey of GSM Wireless Communication System, International Conference on Computer and Information Application (ICCIA 2010).

[2]. Foram Kamdar, Anubbhav Malhotra and Pritish Mahadik Display Message on Notice Board using GSM ISSN 2231-1297, Volume 3, Number 7 (2013), pp. 827- 832 Research India Publications.

[3]. N. Jagan Mohan Reddy and G.Venkeshwaralu Wireless Electronics Display Board Using GSM Technology, International Journal of Electrical, Electronics and Data Communication, ISSN: 2320-2084.

[4]. Shruthi K., Harsha Chawla, Abhishek Bhaduri" SMART NOTICE BOARD", Department of Electronics and Communication, Manipal Institute of Technology, Manipal University, Karnataka.

[5]. Neetesh Saxena and Narendra S. Chaudhari, EasySMS: A Protocol for End-to-End Secure Transmission of SMS IEEE Transactions on Information Forensics and Security, vol. 9, No. 7, July 2014.

[6]. Yi-Bing Lin, Sok-Ian Sou, and Chao-Liang Luo Transmission Policies for Multi- Segment Short Messages" DOI 10.1109/TVT.2457914.2015.

Intrusion Detection System

Mr. Abhijeet Varma, Mr. Subhesh Singh
Information Technology, Thakur College of Engineering and Technology, Mumbai
abhivarma58@gmail.com,subheshs99@gmail.com

***Abstract*— Intruder detector performs the process of detection of any undesirable person or object that might pose some threat in house or an area and the implementation of some important measures to be performed so as to alert the members of the house or an area and increase the security of that place simultaneously. At times**
If an intrusion can be shown on the monitor in a graphical manner then it will enable the user of the system to understand the security of the area in a much more better manner. There is a rise in the requirements of affordable and effective security solutions. In the near future people will prefer security options that require very security system will find it very easy to monitor the safety of their important environments. Thus, this system will alert the owner about intrusion activities and can also take certain measures for protecting the area.
Keywords – Intrusion detection, security using IOT.

I. OVERVIEW INTRODUCTION:

Intrusion, the act of someone that you don't know, who enters into your area without your permission, is on the rise. A human intrusion detection system is devised to detect an unauthorized entry as, if it enters into the building premises it might pose threat to the security hence it must be denied access and an alert must be generated to the security officials to take necessary actions against the possible vulnerability. The purpose of using Security Systems is to safe guard the area of concern hence it finds many of its applications in commercial areas, residential, and in military areas. One cannot always trust humans for security as they might not be completely reliable. In such cases, this system provides proper detection of intruder and provides security. By using this system, we can reduce robbery by detecting the intruder. So we can respond quickly such that no harm takes place in our home or the area of concern. The action that has to be taken would depend on the area of use, like we can have a simple alarm if the intruder enters a residential area. Now a days there is less dependency on the humans for security as they might not be as efficient as security systems and hence they cannot be always relied.

If an intrusion can be shown on the monitor in a graphical manner then it will enable the user of the system to understand the security of the area in a much more better manner. There is a rise in the requirements of affordable and effective security solutions. In the near future people will prefer security options that require very less maintenance. Using the Internet of things technology in the domain of Security will enhance its overall effectiveness and efficiency to a great extent. Using the internet one can monitor, maintain, or modify the security system. Therefore it will enable the user to have complete access to the features and capabilities of the security system and this can be achieved independent of the location of the user that is by using internet.

II. IMPORTANCE OF THE PROJECT:

• Intrusion detection system using Radar can be used in small scale security applications.
• The applications where a cheap security is needed is the biggest market as the system is quite cheap and affordable.
• It can help achieve security for us and the security personnel can see a GUI based RADAR on their screen which will make the managing of the security easier and more user friendly.

III. PROJECT MOTIVATION:

In this present scenario, we can observe that the security systems that we find in the market are quite expensive and there is a big scope for a cheaper security system. We also needed to make a very user friendly application for security, we have hence made a RADAR which enables the user to detect where exactly the intruder is. The user can see the distance and the degrees of the location of the intruder. There are also many restricted places which cannot be monitored at all times hence there we can deploy a cheap and cost effective monitoring system and once the intruder is detected the security officials can see the location of the intruder on their monitoring screens.

IV. PROBLEM DEFINITION:

Theft and Robbery are some of the dangerous threats across the globe in recent times. The home security systems have been constantly failing to safeguard houses and offices. According to a survey conducted by FBI, all the intrusion problems are being occurred in absence of the owners in the home. In the year 2012, a house is being invaded by the intruders for every twenty seconds. Among these, only 10% of them end up in arrests, and rest of them are still suffering from intruders. There are many solutions for this problem like home automation systems, alarm detection systems etc. but there is a chance of producing false alarms by these Systems.

V. LITERATURE SURVEY

IoT Based Human Intrusion Detection System using Lab View

Now a days there is a rise in thefts. This creates a very bad environment for people to live in fear. The issue with home security in the today's world is a cause for concern. The traditional intruder detection system now we are using are very expensive and there can be a possibility of false alarms. This issue can be solved by building a home intruder detection system that can precisely detect a human intruder,

while filtering out movements that are caused due to any other moving objects using LabVIEW and Python. [1]

Internet of Things for Smart Cities

The purpose of this paper is to discuss a general reference framework for the design of an urban IoT. It describes the specific characteristics of an urban IoT, and the services that may drive the adoption of urban IoT by local governments. [2]

The Future Internet of Things: Secure, Efficient and Model-Based

This paper proposes a solution for the possible problems regarding privacy and security. It presents an Architecture applying process knowledge to provide protection through abstraction and privacy through remote data fusion. [3]

VI. FEATURES OF THE PROJECT:

1. Simple and easy to install and configure.

The project contains simple components which do not require complex understanding and can be very easily deployed by the users of this project. There is one ultrasonic sensor and one servo motor that is used also one buzzer has been used to generate an alert.

2. Saving energy and resources, so that it can be utilized in proper way and amount.

3. Automating many of the tasks in the security may help to improve the safety.

One does not need to continuously manage the system as it automatically scans the area and alerts the user if an intrusion is detected.

4. Detection of intruders in real time by using buzzer and radar.

The use of buzzer is simple yet effective as it can signal the intrusion very effectively, the Radar GUI used shows the exact position of the intruder so if the security personnel wants to catch the intruder they exactly know where to go.

5. Automated Intrusion Detection System uses ultrasonic sensors to help security personnel maintain safety.

VII. ANALYSIS AND PLANNING

COMPONENTS	**COST (Rs.)**
Arduino uno	345
Ultrasonic sensor x 1	100
Jumper wires	100
Servo Motors ×2	240
Active Piezo buzzer x 1	60
TOTAL	845

Table 3.1 Cost

HARDWARE:

Arduino UNO ×1:

Fig. 3.1 Arduino UNO

The Arduino Uno is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc. This board consists of sets of digital and analog input/output pins that may be interfaced to various expansion boards and other circuits.

Servo Motors x 1:

Fig. 3.2 Servo motors

Servo motors are DC motors that allows for specific control of angular position. The servo motors are actuallyDC motors whose speed is gradually lowered by the gears. The servo motors usually have a revolution cut- off from 90° to 180°. A few servo motors also have revolution cut-off of 360° or more. But servo motors do not rotate constantly. The rotation of the servo motor is restricted in between the fixed angles

Jumper wires:

Active Piezo Buzzer:

Fig.3.5 Piezo buzzer

An active piezo buzzer will generate a tone using a internal oscillator, so all that is a DC voltage. It can be used as an alarm for indicating an event in a project.

Software:

Fig. 3.3 jumper wires

The jumper wires are used to transfer electrical signals from one point to another point in the circuit. The wires are used for modifying the circuits or to detect problems within a circuit. Jumper wires can have various colors and sizes depending on the area of application. Jump wires are used in breadboards to set up connections between the central micro controller and other devices such as buttons and sensors.

Breadboard:

Fig. 3.4 Breadboard

A breadboard is a construction base for prototyping of electronics. In the 1970s the solder less breadboard became available, nowadays the term "breadboard" is commonly used to refer to these.

Arduino IDE:

It is software package used to program microcontrollers by writing a program in high level Arduino language which is then compiled to binary and uploaded to development board at correct baud rate depending on the board itself.

Circuit Diagram:

Fig. 3.6 Circuit DiagramDESIGN OF PROJECT:

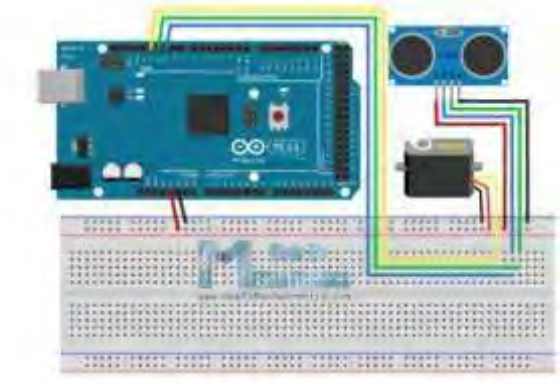


Fig. 3.7 Design of Project

VIII. FLOWCHART

many people will get attracted towards this system. For future scope we can add more functionalities, to make a more holistic device and increase its acceptance.

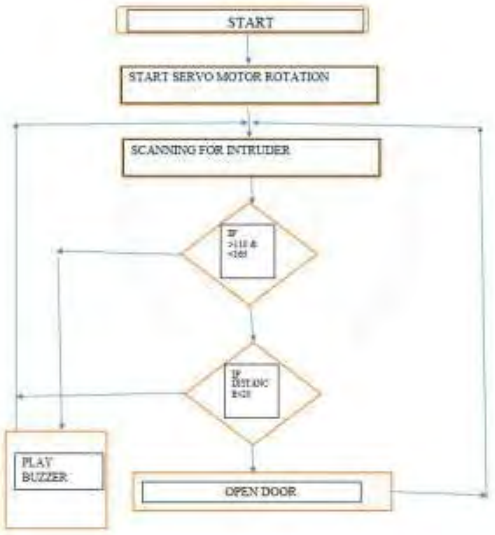
Its future scope is quite vast. The project made by us is a simple prototype and an actual model can be made using simple ultrasonic sensor and a servo motor. Many features can be added to the existing project like enabling the user to

be able to view the current status of the concerned area on a mobile phone. Another feature that can be added is generating SMS to the user and also Email may be send on the users' device.

One can also add features such as a dashboard online on a website which would show the intrusion attempts that were made in the past week, past month, past day, or any time period specified by the user of the security system. By analyzing all these parameters one can improve their security furthermore by seeing its severity.

IX. V. RESULT

We have developed a device that can detect intrusion of any person in restricted place. If someone trespasses then he will be alerted using a buzzer. A display show the location of the person, distance that how long he/she is and at what degree. If a person detected than radar becomes red indicating presense of a person. We have Lessen the cost of this security device, so that it can be made available on a small scale.



X. WORKING:

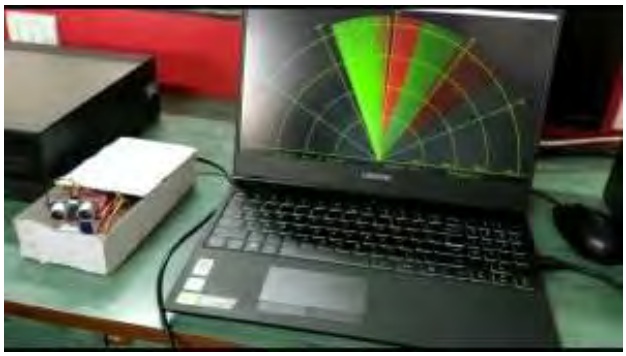


Fig. 3.8 Flowchart

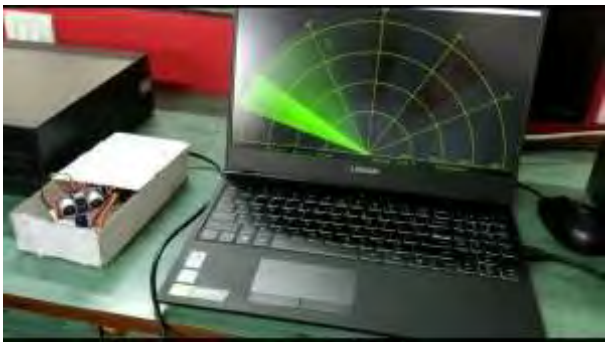


Fig 5.1 Actual Working model

The Arduino device needs to be connected to the laptop device where the GUI of the RADAR has to be seen. After this the Ultrasonic sensor that is attached to the Servo motor will start scanning the surrounding area. This can be seen on the Laptop. If there is any intruder in the given range then it will be marked by red on the RADAR GUI. If the intruder enters the region which sensitive then the buzzer will get activated. When the intruder goes out of boundary then the buzzer will automatically get triggered off.

XI. SCOPE OF THE PROJECT :

This project can have application where a simple and a cost effective security is needed. If the security is busy in some other task then their system will detect and alert any intruder in restricted area. As it is quite affordable

XII. CONCLUSION

Generally, these days' people make use of CCTV cameras for security of their home or offices but it has a few disadvantages like, it can simply record the area and if theft takes place you can see the recording for any information related to the trespasser, but we want a mechanism or arrangement to alert the owner about the threat so that it can take immediate action or some predefined action can be carried out. Therefore we have created our simple and affordable solution for the problem of intrusion in an area of concern. Now with our human intrusion detection system we are able to keep a constant watch over the house. As soon as the system detects any intervention, it will start buzzing and will alert the surrounding. We won't require any memory storage because we are not functioning like a CCTV camera. As this system does not contain any expensive components, this system is affordable and can be used at places where a cost effective security solution is desired.

References

[1] IoT Based Human Intrusion Detection System using Lab View Kalathiripi Rambabu, V.Haritha, S.Nikhil Srinivas, P.Sanjana Reddy

[2] Zanella, A., Bui, N., Castellani, A., Vangelista, L. and Zorzi, M., 2014. Internet of things for smart cities. IEEE Internet of Things journal, 1(1), pp.22-32.

[3] The Future Internet of Things: Secure, Efficient, and Model-Based. Joshua E. Siegel ; Sumeet Kumar ; Sanjay E. Sarma. Publication Year: 2018, Page(s): 2386 - 2398

[4] https://howtomechatronics.com/projects/arduino-radar-project/

Hand Gesture Controlled Wheelchair

Anjali Chauhan, Pradum Chauhan, Vaibhav Dixit (Te-Ita-21)
Thakur College of Engineering and Technology, Mumbai, India
chauhananjali8991@gmail.com, chauhanpradum1681@gmail.com, vaibhavdixit313@gmail.com

Abstract—**Research suggests that an estimated 1% of the world's population, or just over 65 million people, need a wheelchair. Joystick-controlled wheelchairs are available on the market these days whose cost range ranges from Rs 60,000 to Rs 90,000. This Hand Gesture Controlled wheelchair will be prepared in less than Rs 15,000. This paper is to design a Hand Gesture Controlled Wheelchair for the bodily disabled people who face difficulty moving in today's life from one place to another. This paper is intended to build a wheelchair helpful to the physically handicapped person which can be controlled by hand gesture by using accelerometer. This uses a mechanism that controls the user's hand gestures in the wheelchair and interprets the user's motion and moves accordingly. The user can use simple hand gestures to move the wheelchair without any external help or support. This can be used in both hands and can be operated from a distance to reach the user.**

Keywords – IoT, Arduino UNO, gesture control, Accelerometer, transmitter, receiver, wheelchair

I. INTRODUCTION

The term IoT refers to things we use on a daily basis that also connect to the Internet, allowing us to monitor or receive data about the "thing" from our smartphone or computer. Internet of things helps us to work smarter, live smarter and deal with our lives, but it also supports behind the scenes wellbeing. Research says that One billion individual, or 15% of the world's population, experience some form of disability, and handicap commonness is higher for developing countries. It is therefore extremely important to build a system in order to make their lives easier and more versatile. Most people now need a wheelchair because of the growing percentage of elderly people and people with disabilities who want to make their personal mobility easier.

This article introduces a hand gesture-based wheelchair system for handicapped person, who are partially paralyzed. In this process the physical movement of the hand is sensed with sensors and accordingly the wheelchair is controlled. This paper is to demonstrate that accelerometer can be used to control wheelchair using simple hand gestures. The purpose of this project is to control a wheelchair using an accelerometer and RF module. The system is enabled and working based on user hand gestures. Accelerometer is a profoundly sensitive sensor and capable of detecting the tilt. It senses the tilt and the signal is transmitted to the receiver via transmitter; interfaced with Arduino and accelerometer. The sensed signal is transformed into electrical signal and then the signal is transmitted after converting analog signal into digital signal. If the tilt is on the right side then the wheelchair moves to right direction or to left if the tilt is to the left side. Wheelchair movement can be controlled in Forward, Right, backward and left direction by using accelerometer and RF module transmitter and receiver.

Most of the times, disabled people find difficult when they need to move from one place to other place hence, depends on others to move from one location to other location. The wireless wheelchair is designed in such a way that it can be operated by the user by using simple hand gestures. This can be used in both hands and can be operated from a distance to reach the user. The wheelchair is designed in such a way that commands can be sensed and transmitted within the range of 20 meters.

The gesture-based wheelchair is suitable for the elderly and the physically challenged individual who, due to paralysis, birth or old age, are unfortunate to have lost mobility in their limbs. Elders find it difficult to travel to travel inside the house for day to day activities without support or external aid. The hand gesture control wheelchair can be used by older or physically challenged move inside or out the home without trouble an outside guide. The low cost design is one of the main concerns to develop this wheelchair.

II. SYSTEM DESIGN

The wheelchair system design and the required components are discussed in this section. The wheelchair system contains two parts, they are transmitter and receiver. The block diagrams of transmitter and receiver are shown in figure A and figure B respectively. The components are classified into two types, one is hardware and the second one is software component.

A. Hardware Components:

Arduino UNO: The Arduino UNO is an open-source microcontroller module based on the Microchip AT mega 328P microcontroller. The board is fitted with sets of digital and analog input/output pins that can be interfaced with different boards and other circuits. The board has 14 Digital pins, 6 Analog pins, and can be programmed using a type B USB cable with the Arduino IDE.

Accelerometer: The MPU6050 is a Micro Electro-Mechanical Systems that contains a 3-axis Gyroscope inside it. It allows us to calculate a device or object's acceleration, velocity, orientation, displacement and many other parameters related to motion. This module is also fitted with a Digital Motion Processor (DMP) which is powerful enough to perform complex calculations and free up the work for Microcontroller. The accelerometer is interfaced with transmitter of RF module and Arduino UNO.

RF Module: An RF module (radio-frequency module) is a small electronic unit used between two devices for transmitting and/or receiving radio signals. The transmitter and receiver are interfaced via RF module. The RF module receiver and communicate with another wirelessly.

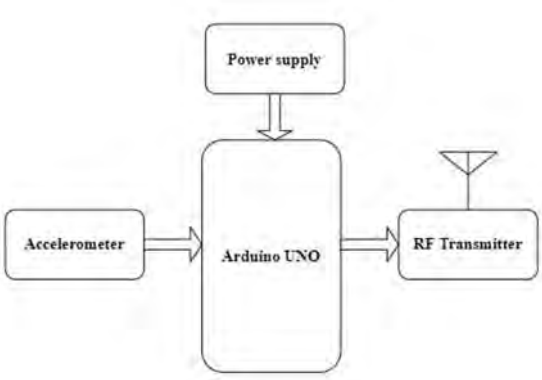


Figure A Block Diagram of Transmitter

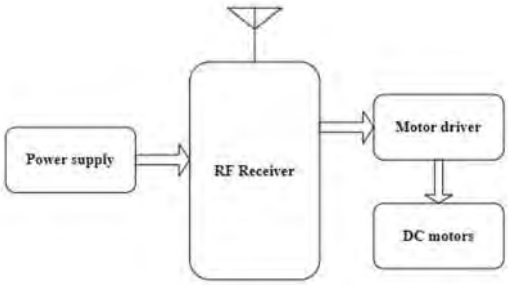


Figure B Block Diagram of Receiver

DC Motor driver: The DC motor-driver (L293D) controls the DC motor in possible rotations. In the present work the motor-driver controls two DC motors simultaneously. That means it controls the direction of two motors simultaneously which will help in moving vehicle wheels synchronously.

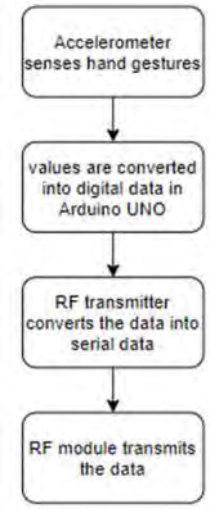
B. Software Component:

Arduino IDE: Arduino IDE is an open source software primarily used to compile and write the code into the Arduino Module. The IDE environment mainly includes two basic parts: Editor and Compiler where the former is used to write the necessary code and later to compile and upload the code into the provided Arduino Module.

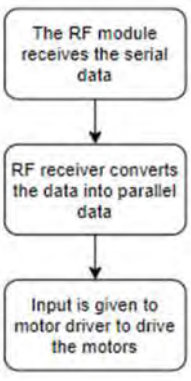
III. IMPLEMENTATION

The hand gesture controlled wheelchair is designed by using Arduino UNO, RF module and accelerometer. An accelerometer sensor is used to detect different hand gestures to move the wheelchair. In the present work MPU6050 is used as accelerometer sensor. The sensor can be either fixed directly with the hands of the person or can wear with hand gloves. Both MPU6050 accelerometer and Arduino UNO is connected to the transmitter of RF module which enables to move the wheelchair with some simple hand gestures. The hand gestures are sensed by the accelerometer and transmitted to the RF receiver. RF receiver receives the signal from MPU6050 via transmitter. RF receiver receives the signal from MPU6050 via transmitter. When the signal is received to the receiver form the transmitter, it triggers the Motor driver which results in movement of the wheelchair.

Data flow diagram:



a) Transmitter side



b) Receiver side

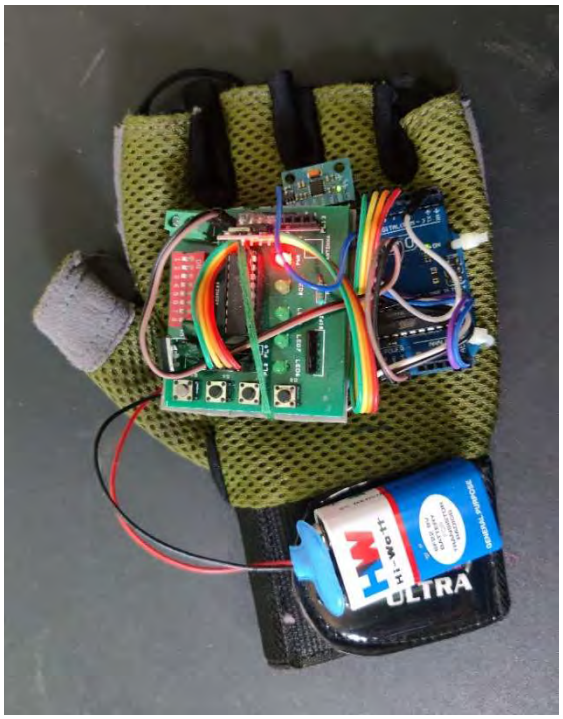


Figure C Prototype of Transmitter

Figure D Prototype of receive

Figure E Prototype of model design

IV. RESULT

The functional operations are implemented in the Arduino IDE software. The action of various commands used in the program is shown in table 1. When the palm movement is upward and downward, wheelchair moves backward and forward respectively. When the hand gesture is towards left or right the wheelchair moves towards left or right direction respectively. When the palm is set to horizontal, wheelchair stops.

Table 1 Functional Table

Palm Movement	Action
Upward	Wheelchair moves backward
Downward	Wheelchair moves forward
Left	Wheelchair moves left
Right	Wheelchair moves right
Horizontal	Wheelchair stops

V. FUTURE SCOPE

- Along with hand gestures, joystick can also be used to control the wheelchair.
- The wheelchair can be controlled by using smartphone.
- Voice commands IC's can be used to control the movement of the wheelchair.
- The wheelchair can also be controlled by using EEG brain waves.
- In future the wheelchair can be operated by using hand and head gestures, joystick, smartphones and brain signals, all in one.

VI. CONCLUSION

This paper presents a hand gesture controlled wheelchair with many benefits such as reduced complexity, simple control, low cost and high reliability compared to other traditional wheelchairs. The gesture controlled wheelchair is designed by using Arduino UNO, RF module and accelerometer. The wireless wheelchair is designed in such a way that it can be operated by the user by using simple hand gestures as shown in table 1. The gesture operated wheelchair can be used without trouble and without external help by elderly or physically disabled moving inside or out the house.

VII. REFERENCES

[1] Saud Landge, Anupmani Tripathi, Nilesh Naik, Vineet Yadhav, Archana Chaugule and Jaymala Chavan," Accelerometer based gesture control the wheelchair", International journal of Electronics, Electrical and Computational System, Vol.5, Issue 11,November 2017,ISSN2348-117X.

[2] S. Manogna1, Sree Vaishnavi2, B.Geethanjali "Head movement based assist system for physically challenged".

[3] Swarna Prabha Jena, Sworaj Kumar Nayak, Saroj Kumar Sahoo, Sibu Ranjan Sahoo, Sarswata Dash, Sunil Kumar Sahoo, "Accelerometer Based Gesture Controlled Robot Using Arduino", International journal of Electrical and Computational System, Vol.3, Issue 1,April 2018,ISSN4785.

[4] Prof. Vishal V. Pande, Nikita S. Ubale, Darshana P. Masurkar, Nikita R. Ingole, Pragati P. Mane," Hand Gesture Based Wheelchair Movement Control for Disabled person Using MEMS", Journal of Computer Science and Computational System, Vol.16, Issue 1,April 2019,ISSN2685.

[5] Riyaz Mansuri, Sandesh Vakale, Ashish Shinde, Tanveer Patel, "Hand Gesture Control Robot Vechicle", Vol.16, Issue 1,August 2019.

[6] V.Rajesh, "SEMG based human-machine interface for controlling a wheelchair by using ANN", proceedings of Control Automation Communication and Energy Conservation 2009. INCACEC 2009.2009 International Conference.

[7] Mr. Vijendra P. Meshram, Ms. Pooja A. Rajurkar, Ms. Mohini M. Bhiogade, Ms. Arundhati C. Kharabe, Mr. Dhiraj Banewar," Wheelchair Automation Using Head Gesture", Computer Science and Computational System, Vol.6, Issue 1,June 2019,ISSN2.

Arduino Based Hand Gesture Control Of Your Computer

Parth Ranawat, Aditya Salvi, Simran Saroj, Amey Sawant
TCET, Mumbai

Abstract - **With progression of Automation innovation, life is getting less complex and simpler in all perspectives. In this day and age Automatic frameworks are being favored over manual framework. With the fast increment in the quantity of clients of web over the previous decade has made Internet an a vital part of life, and IoT is the most recent and developing web innovation. Internet of things is a developing system of regular item from modern machine to shopper merchandise that can share data and complete undertakings while you are occupied with different exercises. Secure and brilliant home (SSH) utilizing IoT is a framework that utilizes cell phones to control fundamental home capacities and highlights naturally through web utilizing voice control.**
The segments utilized are LED Indicators, Switches, Touch Screens and LCD Displays as a piece of HMI gadgets to speak with PC or robots. Another approach to speak with machines like Robots or Computers is with the assistance of Hand Gestures.Instead of utilizing a console, mouse or joystick, we can utilize our hand signals to control certain elements of a PC like play/stop a video, move left/right in a photograph slide appear, look up/down in a website page and some more.
Arduino based hand signal control where you can control hardly any elements of your internet browser like exchanging between tabs, looking all over in site pages, move between errands (applications), play or interruption a video and increment or reduction the volume (in VLC Player) with the assistance of hand motions.
Keywords – LED indicators,Touch screen,Arduino Nano, UltraSonic Sensors, Jumper wires etc.

I. INTRODUCTION

The internet of things, or IOT, is an arrangement of interrelated figuring gadgets, mechanical and advanced machines, items, creatures or individuals that are given extraordinary identifiers (UIDs) and the capacity to move information over a system without expecting human-to-human or human-to-PC communication.

A thing in the internet of things can be an individual with a heart screen embed, a livestock with a biochip transponder, a vehicle that has worked in sensors to alarm the driver when tire pressure is low or some other characteristic or man-made item that can be relegated an IP address and can move information over a system

Literature Survey

1Real –Time natural Hand Gestures,2005-B.Yi, F.C. Harris Jr., L. Wang and Y Yanapplied multivariant Gaussian dispersion to perceive hand signal utilizing non-geometric features.The input hand picture is portioned utilizing two diverse methods,skin shading based division by applying HSV shading model and bunching based thresholding techniques.Some tasks are performed to catch the state of the hand to extricate hand feature.The altered Direction Analysis Algorithm are received to discover a connection between statical parameters from information and use to figure object incline and pattern by finding the bearing of the hand motion

2 Apparatus and Method for Gesture Recognition, 2012-Pratik Bhatt, Deval Mehta, Nirav Faraswamiproposed a collaborating method for constant following of hand catching motions through a web camera and Intel Pentium based individual computer.This framework is actualized without theutilization of modern picture handling calculation and hardware.It utilizes coin measured accelerometer sensor for controlling the application.Hand signal procedures is done progressively condition utilizing vision based systems.

3Automated Mobility and Orientation System for Blind or Partially Sighted People, 2013-Abdel Ilah Nour Alshbatatdesigned an assistive innovation gadget called the electronic glove to fill in as a portability help for daze and outwardly disabled individuals. The creator executes the glove with an ergonomic plan and an installed electronic framework, which fits inside the handle of the glove. The framework was structured utilizing hepatic sensors to identify the impediment above thewaistline. It works in such a manner when the obstructions isdetected the glove vibrates or makes a sound. This framework just recognizes deterrents over the waistline.

4Hand Gesture Recognition System using camera, 2014-In 2014, Viraj Sinde, Tushar Bacchav, Jitendra Pawar and Mangesh Sanap created "Hand Gesture acknowledgment framework utilizing camera" they center around utilizing guiding practices for a characteristic interface toward group the dynamic hand motion, they built up a straightforward and quick movement history picture based framework. This paper shows low multifaceted nature and motion acknowledgment unpredictability and progressively appropriate for controlling constant PC framework. It is material just for the use of intensity point introduction

5Remotely controlled Power point presentation navigation using hand gesture, 2012-In 2012, Ram Rajesh J, Nagarajunan D and Aarthi R, "Remotely controlled force point introduction route utilizing hand motions" built up the framework which slides of intensity point introduction without utilizing any marker and gloves. In the wake of

distinguishing hand estimation is utilized for hand figures. On the off chance that the figures are not extended appropriately while making a signal, at that point application doesn't work appropriately.

Requirement Specification

Useful Requirements**:** Functional necessity are the capacities or highlights that must be remembered for any framework to fulfill the business needs and be satisfactory to the clients. In light of this, the practical prerequisites that the framework must require are as per the following: The framework ought to have the option to anticipate a rough offer cost. The framework should gather exact information from the site in predictable way.

Non-Functional Requirements**:** Non-practical necessity is a portrayal of highlights, qualities and trait of the framework just as any limitations that may restrain the limits of the proposed framework. The non-useful necessities are basically founded on the exhibition, data, economy, control and security proficiency and administrations. In light of these the non-practical prerequisites are as per the following:

1. The framework ought to give better precision.

2. The framework ought to have straightforward interface for clients to utilize.

3.To perform productively in short measure of time.
Methodology
The Arduino ought to be modified to peruse the separation of hand from the US sensor. The total program is given toward the finish of this page; just beneath I have given the clarification for the program. On the off chance that you are new to Ultrasonic sensor, simply experience Arduino and Ultrasonic Sensor Based Distance Measurement. By perusing the estimation of separation we can land at specific activities to be controlled with signals, for instance right now have customized 5 activities as a demo.

Step 1: When both the hands are put up before the sensor at a specific far separation then the video in VLC player should Play/Pause.

Step 2: When right hand is set up before the sensor at a specific far separation then the video should Fast Forward one stage.

Step 3: When left hand is set up before the sensor at a specific far separation then the video ought to
Rewind one stage.

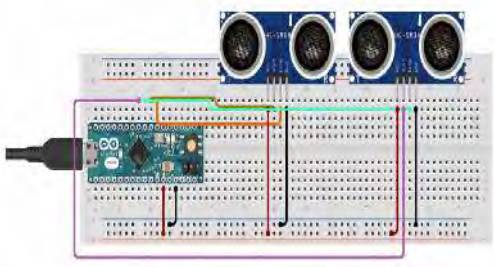
Step 4: When right hand is set up before the sensor at a specific close to separation and afterward whenever moved towards the sensor the video should quick
advance and whenever moved away the video ought to Rewind.

Step 5: When left hand is put up before the sensor at
a specific close to separation and afterward whenever moved towards the sensor the volume of video
should increment and whenever moved away the
volume should Decrease.
Technology

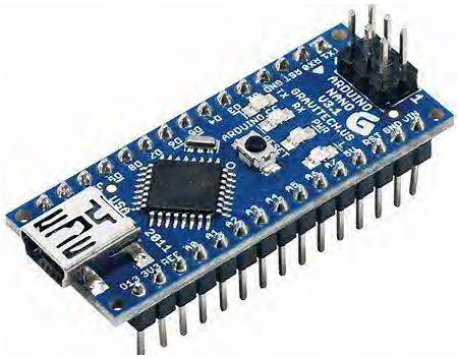
The source code for the IDE is discharged under the GNU General Public License, form 2. The Arduino IDE bolsters the dialects C and C++ utilizing extraordinary principles of code organizing. The Arduino IDE supplies a product library from the Wiring venture, which gives numerous basic info and yield strategies. Client composed code just requires two fundamental capacities, for beginning the sketch and the principle program circle, that are arranged and connected with a program stub primary() into an executable cyclic official program with the GNU toolchain, likewise included with the IDE appropriation. The Arduino IDE utilizes the program avrdude to change over the executable code into a book document in hexadecimal encoding that is stacked into the Arduino board by a loader program in the board's firmware

II. DESIGN PHASE

Circuit diagram



Software



Arduino Nano

UltraSonic

Jumper wires

1. Arduino IDE:

It is programming bundle used to program microcontrollers by composing a program in elevated level Arduino language which is then incorporated to double and transferred to improvement board at right baud rate contingent upon the board itself.

2. Python

Python is a mainstream programming language. It was made in 1991 by Guido van Rossum.

It is utilized for:
Hardware
eb development (server-side)

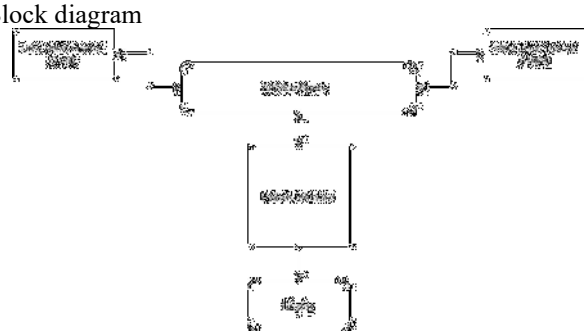
- software development

III. WORKING PRINCIPLE OF THE SYSTEM

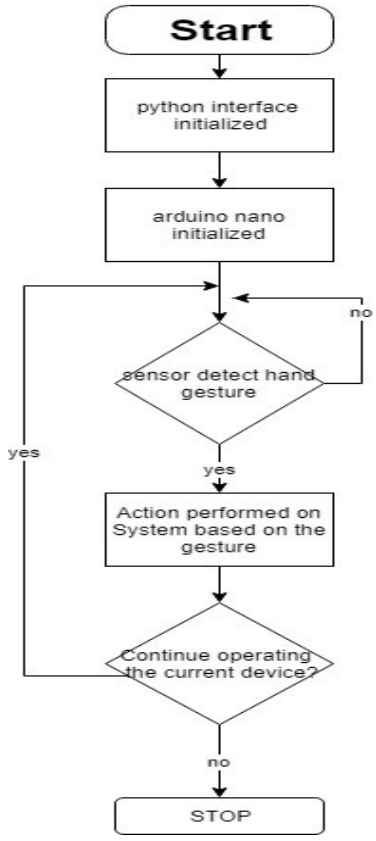
Algorithm
Step 1**:** Open Windows Command prompt and change the directory to the folder where you have installed python.
Step 2**:** Inside your python directory use the command *python –m pip install –upgrade pip* to upgrade your pip. Pip is a tool in python which helps us to install python modules easily. Once this module is upgraded (as shown in picture below) proceed to next step.
Step 3**:** Use the command "*python –m pip install pyautogui*" to install the pyautogui module. Once the process is successful you should see a screen something similar to this below.
Step 4: Now run the code on IDE.
Step 5: Make connections as per circuit diagram.

Step 6:Perform hand gestures and control your machine.

Block diagram



Flow chart



IV. RESULT AND OUTCOME

User can deal with application from separation without utilizing console and mouse by this strategy. This application gives a novel human PC interface by which a client can control media player (VLC) utilizing hand motion. The application characterizes some motion for controlling the elements of VLC player. The client will give signal as a contribution as indicated by intrigued work

V. CONCLUSION

In current world numerous offices are accessible for giving contribution to any application a few needs physical touch and some without utilizing physical touch (discourse, hand signal etc.).But relatively few applications are accessible which are controlled utilizing present and savvy office of giving info which is by hand motion .By this strategy client can deal with application from separation without utilizing console and mouse. This application gives a novel human PC interface by which a client can control media player (VLC) utilizing hand motion. The application characterizes some signal for controlling the elements of VLC player. The client will give motion as a contribution as indicated by intrigued work. The applicationprovides an adaptability of characterizing client intrigue signals for explicit order which make the application progressively helpful for genuinely tested
individuals, as they can characterize the motion as per their achievability.

References

[1]. B. Yi, F.C. Harris Jr., L. Wang and Y, Yan, "Real – Time natural Hand Gestures", *In proceeding of IEEE computing in Science and Engineering,*Jan 2005

[2] Pratik Bhatt, Deval Mehta, Nirav Faraswami, "Apparatus and Method for Gesture Recognition",Gujarat Technological University, May 2012

[3] Abdel Ilah Nour Alshbatat, "Automated Mobility and Orientation System for Blind or Partially Sighted People",*International Journal on Smart Sensing and Intelligent System,* vol.6, no. 2, April 2013

[4] Viraj Shinde, Tushar Bacchav, Pawar and Mangesh Sanap "Hand Gesture Recognition System using camera", *International Journal of Engineering Research &Technology (IJERT),*Vol.3 ,issue 1,January -2014.

[5] Ram Rajesh j, Nagarjunan D, and Aarthi R. "Remotely controlled Power point presentation navigation using hand gesture", *Proc. of the Intl Conference on Advances in Computer,*Electronics and Electrical Engineering, 2012. Conclusion

Arduino Based Smart Blind Stick

Sandeep Pandey, Prem Patel, Riddhi Patel
Information Technology, Thakur College Of Engineering and Technology, Mumbai, India
Sp348220@gmail.com, patelprem616@gmail.com, patelriddhi1519@gmail.com

Abstract— **In the present scenario, a visually impaired individual is constrained to depend on someone else and can't make a trip autonomously to wherever without the assistance of others. So, here we propose an smart blind stick that allows visually challenged people to navigate with ease using advanced technology. The blind stick is integrated with ultrasonic sensor, soil moisture sensor along with tilt sensor. On detecting obstacle, the sensor passes this information to the microcontroller. The microcontroller then processes this information and figures if it is sufficiently close If the obstacle is close the microcontroller sends a signal to sound a buzzer. If the obstacle isn't that nearby the circuit sits idle. It also detects water or mud and sounds a buzzer in different pattern and alerts the blind. By using tilt sensor blind man could remotely locate his stick. We have also used a motor vibrator which provides different vibration pattern for alerting the user. Thus, proposed system helps the people who are visually impaired in the aspect of mobility and making it safer by avoiding accidents.**

Keywords – Ultrasonic sensor, Soil Moisture Sensor, Tilt Sensor, Arduino UNO, Smart Blind Stick

I. INTRODUCTION

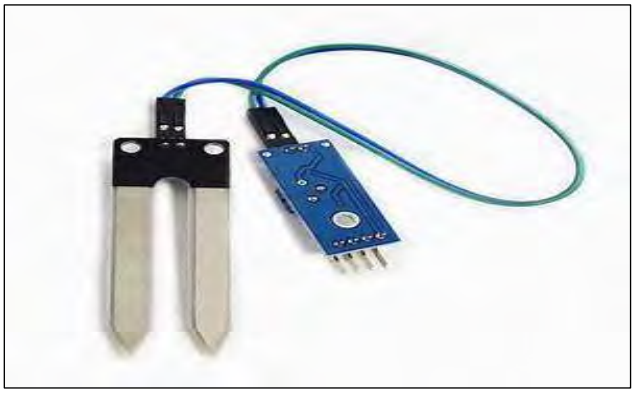
Visual deficiency is a common regular inability among the peoples throughout the world. As per the World Health Organization (WHO) 285 million individuals are outwardly impeded around the world, 39 million are visually impaired and 246 have low vision. About 90% of the world's outwardly impeded live in developing nations. Visually impaired persons have difficulty to interact with their environment. They have a very little contact with surroundings. Physical movement or mobility is a challenge for visually impaired persons, because it can become difficult to recognize obstacles appearing in front of them, and they are not able to move from one place to other. Smart Blind Stick is specially designed to detect obstacles which may help the blind to navigate care-free. The audio messages will keep the user alert and comparatively reduce accidents. The proposed system contains the ultrasonic sensor, Soil moisture sensor, Buzzer and Arduino UNO. The main objective of the paper is to discuss about development work of a stick that could communicate with the users through voice alert and vibration, which is named Smart Blind Stick which involves coding as software and physical installation as hardware part.

II. PROBLEM DEFINTION

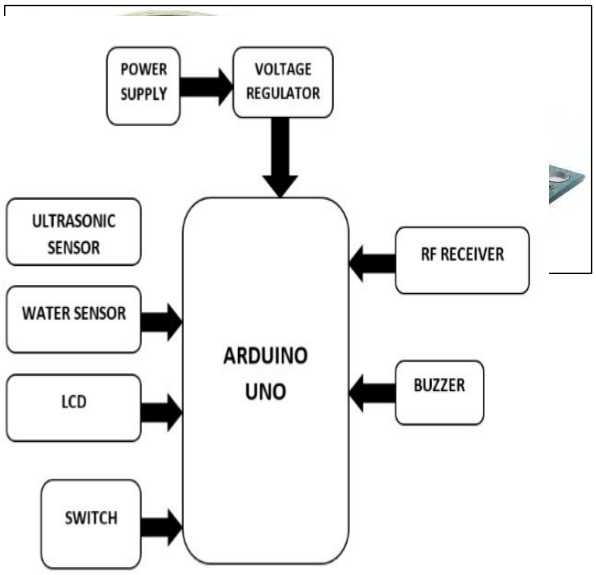
Physical movement is a challenge for visually impaired person. The conventional walking stick used by them is inefficient in detecting obstacle. The disadvantage of conventional cane, However, is its failure to detect obstacles outside of its reach. That is the user has to tap to the ground or obstacle to detect it. The visually challenged people can avoid the object better if the walking stick can produce sound and vibration warnings when there is a object in specified range.

III. BACKGROUND

More than 161 million people worldwide are visually impaired. Among them, 124 million have Low vision and 37 million are blind. Another 153 million people suffer from visual impairment Due to uncorrected refractive errors such as near-sightedness, far-sightedness or astigmatism. Virtually all these people could restore normal vision with eyeglasses or contact lenses. More than 90% of the world's visually impaired people live in low- and middle-income countries. Blindness is a condition of lacking visual perception and it is always described as severe visual impairment with residual vision. The blind people's life and activities are greatly restricted by loss of eyesight. They can only walk in fixed routes that are significant in their lives, with blind navigation equipment's and the accumulated memories in their long-term exploration. This situation has resulted in many difficulties to them in their normal work, lives, activities, and so on. Based on the investigation about daily activity characteristics and modes of the blind, the study found that the main difficulties encountered in a trip of the blind included walking on the road, finding way, taking a bus and looking for usual life-arena Several devices have been developed for mobility and navigation assistance of the blind and are typically known as travel aids or blind mobility aids.

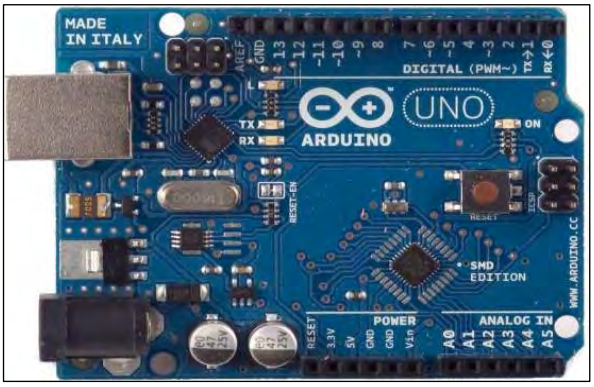


IV. BLOCK DIAGRAM



V. HARDWARE REQUIREMENTS

1. Arduino UNO:



Arduino UNO is a microcontroller board based on ATmega328p. It has 20 pins out of which 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz quartz crystal, a USB connection, a power jack, an ICSP header and a reset button. It contains everything needed to support the microcontroller. The Operation Voltage is 5V simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started. You can tinker with your UNO without worrying too much about doing something wrong.

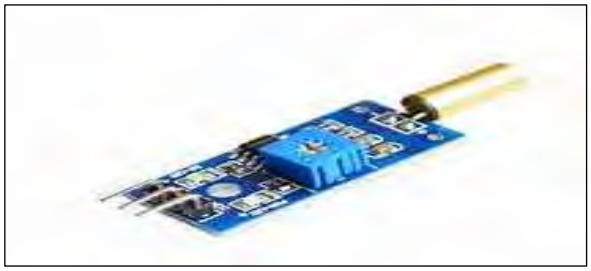
2. Ultrasonic Sensor:

As the name suggest, ultrasonic sensors measure distance by ultrasonic waves. The sensor head emits an ultrasonic wave and receives the wave reflected back from the targeted object. Ultrasonic Sensors measure the distance to the object by measuring the time between the emission and reception. An optical sensor has a transmitter and receiver, whereas an ultrasonic sensor uses a single ultrasonic element for both emission as well as reception. In a reflective model of ultrasonic sensor, a single oscillator emits as well as receives ultrasonic waves alternately.

3. Soil Moisture Sensor:

Soil moisture sensors measure the volumetric water content in soil. The direct measurement of free soil moisture process includes drying, removing, and weighing of a sample.Soil moisture sensors calculates the volumetric water content indirectly by using various property of the soil, such as dielectric constant ,electrical resistance or interaction with neutrons, as a proxy for the moisture content. Soil moisture sensors is used to estimate or calculates volumetric water content. There is a separate class of sensors that measure another property in soils called water potential; these are usually known as soil water potential sensors and which includes gypsum blocks and tensiometers.

4.Tilt Sensor:



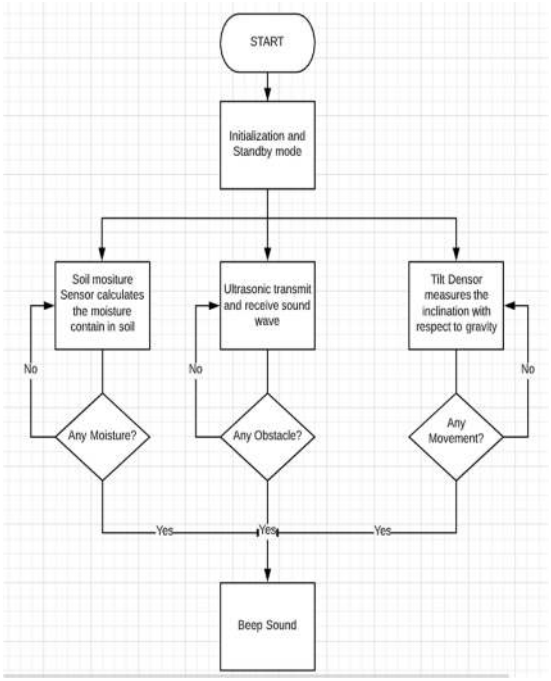
A tilt sensor is an instrument that is used for measuring the tilt in multiple axes with respect to particular plane. Tilt sensors measure the tilting position with respected to gravity and are used in number applications. They enable the easy detection of orientation or inclination of a object. Similar to mercury switches, they also be known as tilt switches or rolling ball sensors. The capability of tilt sensors is influenced by factors together with vibration, gravity, linearity, 0 offset, temperature, cross-axis sensitivity, acceleration/deceleration, shock, clean line of sight between the user and the measured point, and calibration of tilt sensors.

5.Buzzer:

A buzzer is a small efficient component to add sound to our project/system. It is compact and small with 2-pin structure hence can be easily mounted on breadboard, Perf Board and even on PCBs which makes this a widely used component in most electronic applications. There are commonly two types of buzzers available. First one is a simple buzzer which will make a Continuous Beeeeppp.... sound and other type of buzzer is called as readymade buzzer which looks bulkier than simple buzzer and produce a Beep. Beep. Beep. Beep Sound is produced because of the internal oscillating circuit present inside it.

6. Bread Board:

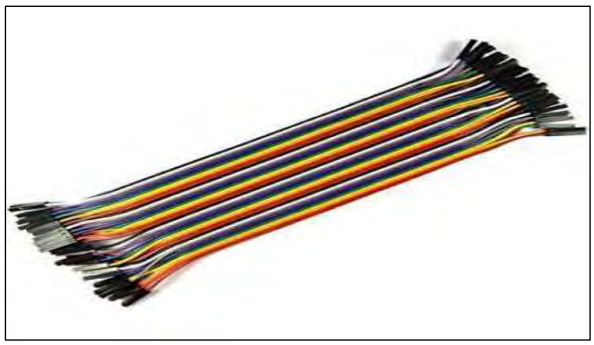
A breadboard is a solderless device for prototype



model with hardware and test circuit structures. Components in electronic circuits can be interconnected by inserting their terminals into the holes and then making required connections through jumper wires. The breadboard has metal strips underneath the board and which connect the holes on the top of the board.

7. Jumper Wires:

A jumper wire is a conducting wire which is used to



transfer electrical signals between two ends in a circuit. The wires can be used to modify circuits as well as to diagnose problems within a circuit. Jumper wires typically vary in size and color depending on what they are being used for. Jumper wires are used to establish connections between the central micro controller and other devices such as sensors and buttons..

VI. SOFTWARE REQUIRMENTS

The Arduino Integrated Development Environment (IDE) is a cross-platform application for Windows, Linux, macOS that is written in functions from C and C++. It is used to write and upload various programs to Arduino compatible boards, also, with the help of third party cores, other vendor developmental boards.

The source code for IDE is released under the GNU General Public License v2. The Arduino IDE supports the languages C++ and C with special rules of code structuring. The Arduino IDE supplies a software library from the Wiring project, which provides many input-output procedures. User defined code requires two basic functions, for starting the sketch and the main program loop, that are compiled and linked with a program into an executable cyclic executive program with the GNU tool-chain, also included with the distribution of IDE. The Arduino IDE provides the program avrdude to convert the executable code into text file in hexadecimal encoding format that is loaded into the Arduino board by a loader program in the firmware of board.

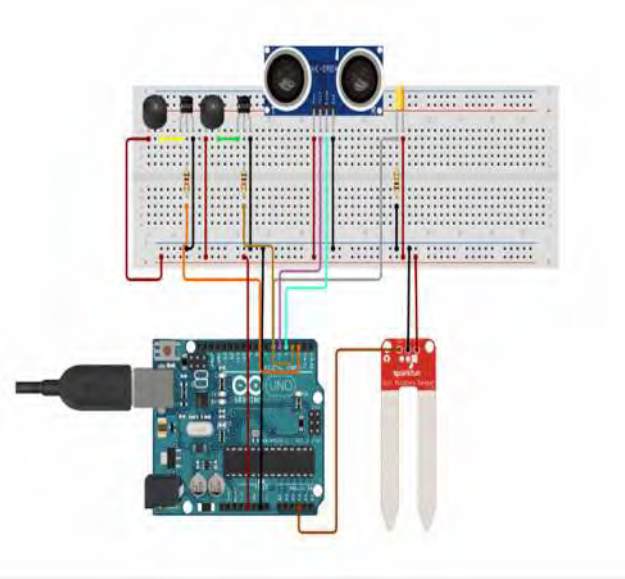
VII. DESIGN AND DEVELOPMENT:

1. Flowchart:

When circuit is switched ON the sensors placed on the stick goes into active state. There are three sensors mounted on the stick which are ultrasonic sensor to detect obstacle in front of the stick, soil moisture sensor to detect water on the ground and tilt sensor which enables the user to locate the stick from ground. Ultrasonic sensor transmit and receive sound wave which in turn helps to detect obstacle. If any obstacle is detected within a specified range then it will produce a beep sound or else do nothing. Soil moisture sensor

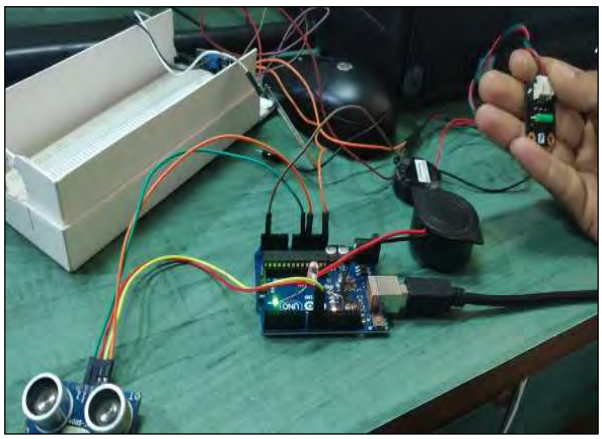
calculates moisture contain in soil.is any moisture contain is detected in the soil then the stick will produce beep sound or else do nothing. Tilt sensor measures the inclination of the stick with respect to gravity. If the stick is inclined then it will produce a beep sound or else do nothing.

2. Circuit Diagram:



VIII. WORKING OF SYSTEM

Smart Blind Stick is basically designed to make sure that the people with visual disabilities can work on their own without being dependent on others for their living. This innovative Stick allows visually challenged people to navigate on their own. This blind stick is integrated with Ultrasonic Sensor so that people with disabilities can walk with a sense of obstacle in between. This ultrasonic sensor makes sure that when there is any obstacle in front of the stick it indicates the user. Ultrasonic sensor senses the obstacle and passes this data to microcontroller. Then this microcontroller will process the data and calculate if the obstacle is close enough or not. If the obstacle is not close enough then it does nothing. If obstacle is close enough then it will produce a buzzing sound so that the user can understand that there is obstacle in front. Buzzer is used so that it can produce a sound when obstacle is ahead. This stick also contains soil moisture sensor. Soil moisture sensor is used so that the user can indicate the path where there is no detection of water and hence selecting a simple path to walk. Soil moisture sensor is connected at the bottom of the stick and when the sensor detects water in the soil then it produces a buzzing sound to indicate the user that the path is not safe. We made use of vibrator motor for the people with hearing as well as visual disabilities. This motor vibrates while buzzing which in turn helps them to detect obstacle in between. This stick includes a tilt sensor which will help to locate the stick on the ground. Tilt sensor is of a great use to the people with visual disabilities. When this stick falls of the ground from the hand of the user then this stick produces a buzzing sound so that the user can easily locate the stick. This tilt Senor measures the angle of the stick with respect to gravity and then detect the inclination to decide whether to produce buzzing sound or not. This stick also contains a LDR which helps to produce light on the stick at times of dark which will allow the people surrounding the user to be aware and hence avoid collation.



This project is designed in such a way that the people with visual disabilities can easily use this stick and be independent. The look of this project is just like a normal stick with some sensors attached which does not give any problem to the user to use this stick.

IX. FUTURE SCOPE

By adding a GPS and GSM module to the stick we can track the actual location of the person as well as establish communication between mobile device or a computing machine. We can also add Vibrator motor which helps to make a vibration by recognizing it the

person holding the stick and also feel the vibration and act accordingly, by this motor this stick can be used by people with hearing disabilities.

X. CONCLUSION

The project proposed a layout, design and architecture of a new concept of Smart Electronic Guiding Stick for blind people. It aims to solve the problems faced by the millions of visually impaired people worldwide in their daily life. The proposed system takes measures to ensure their safety and the main advantage of the system is that it can prove to be very low-cost solution to millions of visually impaired people worldwide with the proposed combination of various working units. It is a real-time system that monitors position of the user and provides dual feedback making navigation more safe and secure than the traditional cane. It can be further improved to have more decision taking capabilities by integrating various types of sensors such as GPS, GSM module and thus could be used for different applications. This our hope is to consider this stick as smart eye for the visual impairments. It is worth mentioning at this point that the aim of this study which is the design and implementation of a smart walking stick for the blind has been successfully achieved.

References

1.https://circuitdigest.com/microcontroller-projects/arduino-smart-blind-stick

2.https://www.instructables.com/id/Smart-Blind-Stick/

3.https://electronicsforu.com/electronics-projects/smart-stick-using-arduino-ultrasonic

4.https://create.arduino.cc/projecthub/hadi1234/arduino-blind-stick-19d865

5. Dambhare, Shruti, and A. Sakhare. "Smart stick for Blind: Obstacle Detection, Artificial vision and Realtime assistance via GPS." In 2nd National Conference on Information and Communication Technology (NCICT).2011; 2:31-33.

6. Al-Shehabi, Monther M., Mustahsan Mir, Abdullah M. Ali, and Ahmed M. Ali. "An Obstacle Detection and Guidance System for Mobility of Visually Impaired in Unfamiliar Indoor Environments." International Journal of Computer and Electrical Engineering. 2014;6(4): 337-341.

7. Jeslin, J. Gnana, A. Vaishnavi, and J. Nivedha. "Enhancement of Smart Cane To Detect And Notify Obstacles.". International Journal of Engineering Research & Technology (IJERT) 2015;4(2):1-3.

Air Quality Detection And Monitoring System

Ayush Jangida, Sarthak Jain, Prakash Jha
Information Technology Thakur College of Engineering & Technology
Mumbai, India
ayushjangida30@gmail.com, sarthakjain3006@gmail.com, prkjha25@gmail.com

Abstract **- There is a surge in pollution of air and sound in recent years. To bring it in check, its monitoring is majorly recommended. To overcome this issue, we are introducing a system through which the the existence of the harmful gases in the surroundings can be detected. The growing pollution at such an alarming rate has started creating trouble for the living beings, may it be high decibels or toxic gases present in the environment, it leaves a harmful effect on human's health and thus needs special attention[1].**

This monitored data can often be obtained from remote locations without actually visiting it, thanks to the access of the internet. The framework of this monitoring system is a combination or collaboration of affective distributed sensing units and information system for data composition. The role of IoT is the new concept that is added in air and sound pollution measurement, which allows data access from remote locations[2].

The level of pollution has increased with time. A lot of factors like the increase in population, increased vehicle use, industrialization and urbanization have harmful effects on human wellbeing by directly affecting the health of the population exposed to it. In order to monitor these effects, in this project we are going to develop an IOT Based Air Pollution Monitoring System through which we will monitor the Air Quality over a web server and mobile application using the internet and will display and analyze the amount of gas present with addition to the temperature and humidity of the given environment. The gases which will be detected are smoke, LPG & CO (carbon monoxide). In this IOT project you can monitor the air quality from anywhere using your computer or mobile[5].

Keywords - Pollution, IoT, Air Quality, Environment, Sensors, NodeMCU, MQ2, DHT11

I. INTRODUCTION

The main objective of IOT Air quality Monitoring System is to highlight the issue of Air and Sound pollution these days. It is necessary to monitor air quality and keep it under check for a better future and healthy lifestyle. Due to flexibility and low cost, the Internet of things (IoT) is getting popular day by day[2].
With urbanization and with an increase in vehicles on road the atmospheric conditions have considerably deteriorated[1]. Harmful effects of pollution include mild allergic reactions such as irritation of the throat, eyes and nose, as well as some acute problems like bronchitis, heart diseases, pneumonia, lung and aggravated asthma.
Monitoring gives measurements of air pollutant concentrations, temperature, and humidity , which can then be analyzed, interpreted and presented. This information can then be applied in many ways. Analysis of monitoring data allows us to assess and measure how bad air pollution is from day to day[4].

II. BACKGROUND

Air quality detection and monitoring has become a major aspect for day to day needs of many professions, from various practical experimentation to detecting the air quality of a specific region and to get the actual observation of air quality.

This system helps to detect the gases present in real time in parts per million in mobile application. It gives the user interface for gauge reading of the gases present and based on some parameters determine whether the air quality is Fresh Air, Poor or Hazardous[6]. However, preceding gas detectors required high voltage power input, further research on power requirement in gas detectors has reduced to make battery power viable making widespread installation of this system in residences highly feasible.

Later developments in gas detectors have sought to improve their performance, reduce power requirement and improve their nuisance alarm sensitivity and also to continuously monitor their status. The most recent advancement in gas detectors have been motivated to make them "smarter"[1]. Gas can be detected either optically (photoelectric) or by physical process (ionization). Detectors may use either or both methods[2].

A - Importance of Project

The requirement of the Air Quality Detection and Monitoring System is because of an increase in the fire prone areas in localized urban regions such as kitchen, buildings, construction region, mining industries petrol pumps. Also to check the real time scenarios of air quality in any other surrounding to determine how good the air is for health and also for further analysis to check the quality of certain regions[6].

- Analysis of certain regions to check temperature, humidity, and other hazardous gases present.

- Display real-time values of the output captured by the sensor.

- Failure to take advantage of this early warning, due to poor performance of the gas detection system, has cost people their lives.

- Easy to transmit data through the internet as the project implements wifi-module[6].

III. TEST AND MEASUREMENTS

A - Testing of hardware

NodeMCUis an open source IoTplatform. It includes firmwarewhich runs on the ESP8266 Wi-FiSoC from Espress if Systems, and hardware which is based on the ESP-12 module. The term "NodeMCU" by default refers to the firmware rather than the development kits. The firmware uses the Luascripting language.Tests will be done by checking the connections everytime and then running a simple test code to blink the inbuilt light to test if everything is working correctly [4].

MQ-2 gas sensor

The Grove - Gas Sensor (MQ2) module is a useful sensor for detecting gas leakage detection (in industry and home). It is also suitable for detecting Smoke,CO,LPG. Due to its high sensitivity and fast response time, measurements can be taken as quickly. The sensitivity of this sensor can be adjusted by using a potentiometer. This should be tested repeatedly by running a test code to blink lights at different levels of Smoke, CO, LPG levels.

DHT 11 Temperature and humidity sensor DHT11 is a low-cost digital sensor for sensing humidity and temperature. This sensor can be easily interfaced with any micro-controller such as Raspberry Pi, Arduino, etc., to measure temperature and humidity instantaneously.

The smoke sensor, DHT11 is highly sensitive to the amount of smoke generated at each point of time in the region. As soon as it senses the smoke above threshold limit, it starts the alert system.

Testing: After the prototype is ready, we first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding it and will check again.

IV. FEATURES AND DESIGN REQUIREMENT

A - Features of the Project

The Project consists of the following features:

- LCD display and chart: All the collected input from sensors (DHT11 & MQ2) are sent to the BLYNK server and thereby the GET request is made and the front-end interface of mobile displays the real time value on mobile application Also on the other hand the data is sent to Thingspeak server where it is analyzed in the form of chart.

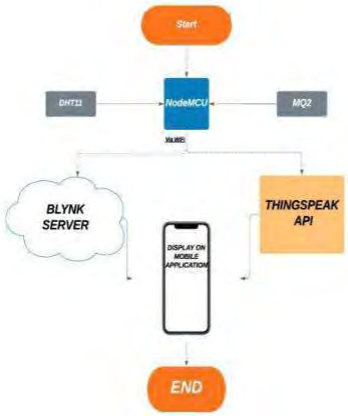
- High sensitivity and fast response: The smoke sensor, DHT11 sensor is highly sensitive to the amount of smoke generated at each point of time in the region. As soon as it senses the smoke above the threshold limit, it starts the alert system.[3]

- Simple drive circuit: The connections and the circuit design of this system is simple hence it is much more convenient to develop.

B - Methodology

The proposed method takes an automatic control action upon detection of gas, temperature and humidity. Our always contains the gases(CO, LPG, Smoke) in some small amount of level, the problem arises when these gases reaches to certain level which harmful to humans and this happens only when there is some sort of leakage detected on the types of gases, whenever the display and analysis of the environment is needed, all that is needed to be done is to plug the nodemcu and the data is sent through Internet to mobile to display and it also displays if the air is fresh, poor or hazardous depending on the ppm level received of LPG,CO & smoke and also the analysis is displayed on thingspeak server[6].

A-Block Diagram



B. Circuit Diagram

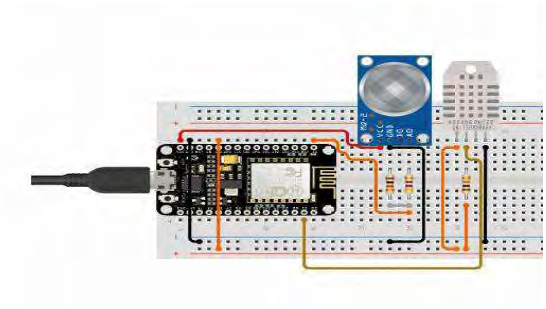


Fig. 2. Circuit Diagram

The above circuit diagram represents the different components and circuit used to connect these components. The Hardware and Software Requirement is discussed later in this paper.

C. GUI Screenshot

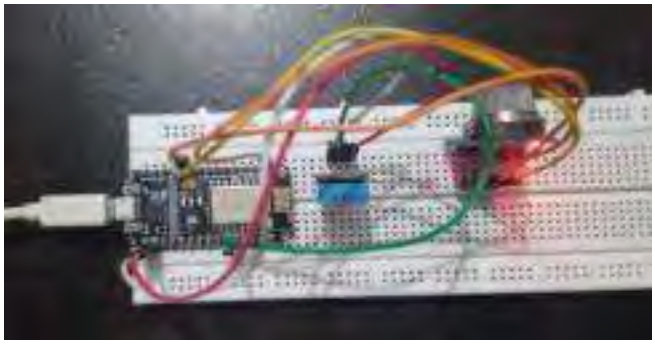


Fig. 1. Block Diagram

The above block diagram represents the actual working of our project. NodeMCU contains inbuilt WiFi, which enables us to send the data captured by the DHT11 and MQ2 sensors to be displayed on Blynk Mobile Application, ThingSpeak Server, and an email notification if level of pollution reaches hazardous level.

Fig. 3. GUI Screenshot

The above image is the actual implementation of Fig. 2.

D. Hardware and Software Tools

TABLE I. HARDWARE AND SOFTWARE TOOLS

Software Tools

E. Feasibility Study

- **Technical Feasibility:** Here one has to test whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible[4].

- **Economic Feasibility:** As part of this, the benefits and costs associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

- **Legal Feasibility:** Legal issues can affect a system's acceptance by users, its performance , or the decisions on whether to use it in the first place. Therefore it is best to consider these explicitly in system design. Clearly, the behavior of those being enrolled and recognized can influence the accuracy and effectiveness of virtually any biometric system.

- **Operational Feasibility:** It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.

- **Social Feasibility:** The acceptability of a biometric system depends on the social and cultural values of the participant populations. A careful analysis and articulation of these issues and their

identification can improve both acceptability and effectiveness.

V. RESULTS

A - Output

- All the levels of output received from DHT11 sensor i.e temperature & humidity and From MQ2 sensor i.e LPG, CO, and Smoke are shown as gauge reading in mobile app and the LCD in mobile displays the quality of air.

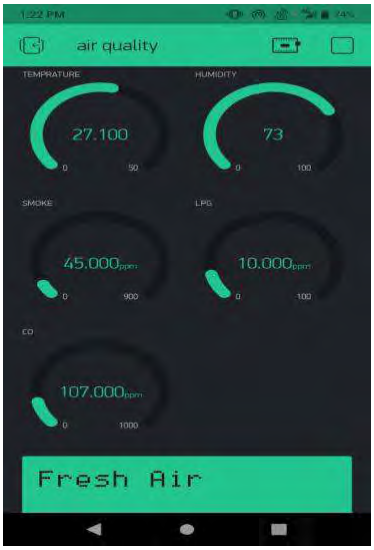


Fig. 4. Blynk Mobile Application

● The second part, that is for the analysis, the data is sent to thingspeak server as POST request and the analysis is displayed in the form of chart.

Fig. 5. ThingSpeak Server

● If the air quality level reaches a hazardous level, then the user gets an email regarding the same that, 'The air quality is currently hazardous'.

Fig. 6. Email from Blynk Application

VI. DISCUSSION

A - Future Scope

● A publicly available website can be implemented and the real time and past analysis of gases, temperature and humidity can be shown of many areas, also after gathering enormous data, Machine Learning can be implemented which can predict the values of those parameters for future and also classify the diseases which can be caused if the gases goes above poor air quality level by classifying the PPM levels

● It can also be used in F1 cars as there is a high possibility of catching fire when a car runs at that speed there is high possibility of them catching fire due to the web based output it can be conveyed to the driver and thus he can save his life and win the race.

B - Conclusion

We have successfully shown the detection and monitoring of Gas, temperature & humidity and by IOT based project is possible which can display and analyze Gases present in a given environment, analyze and monitor them on real time basis using various online tools Thingspeak & Blynk platform also an email is sent to user if the gas reaches to hazardous level, the data here can useful for other organization for further analysis purpose[6]. This application demonstrates the mobile phone's ability to be used as a advanced warning device in the event of a smoke alarm activation, its automated group alert system reduces the chance of accidents to people who have not been able to respond to an alarm.

ACKNOWLEDGMENT

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staff of the IT Department who helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in the laboratory for their timely support.

References

[1] Gabrielle Wong-Parodi, M. Beatrice Dias, Michael Taylor ,"Effect of Using an Indoor *Air Quality Sensor on Perceptions of and Behaviors Toward Air Pollution", Published in JMIR mHealth anduHealth 2018* ,DOI: 10.21.96

[2] Poonam Pal, Ritik Gupta, Sanjana Tiwari, Ashutosh Sharma, "IOT BASED AIR POLLUTION MONITORING SYSTEM USINGARDUINO",e-ISSN: 2395-0056, Volume: 04 , Issue: 10, Oct 2017.

[3] Manish R.Deshmukh,Suraj J. Chavan, Renuka A. Chavan, Prof. Ganesh Atarde, "IOT Based Air & Sound Pollution Monitoring System", International Journal of Advance Research and Innovative Ideas in Education Vol-4 Issue-3 2018, IJARIIE-ISSN(O)-2395-4396

[4] Shifeng Fang, Li Da Xu, Senior Member, IEEE, Yunqiang Zhu, Jiaerheng Ahati, Huan Pei, Jianwu Yan, and Zhihui Liu, "An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things" , *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, VOL. 10, NO. 2, MAY 2014

[5] Ramagiri Rushikesh,Chandra Mohan Reddy Sivappagari,"Development of IoT based Vehicular Pollution Monitoring System",2015 *International Conference on Green Computing and Internet of Things* (ICGCIoT), DOI: 10.110

[6] Dr.Parashuram Baraki, Professor &Head, Dept of CSE, HIT, Nidasoshi. Shankarayya Shastri, Asst Professor, Dept of CSE, SJEC Mangalore. Asif Mohemmed, Dept Of CSE. Anusha Hegde, Asst Professor, "*Real Time Weather Analysis Using ThingSpeak*" Dept of ICT,MIT Manipal. June 11, 2018 ISSN: 1314-3395 (on-line version)

Arduino Based Dual Axis Sun Tracking Solar Panel

Miloni Shah, Swati Solanki, Prakash Upadhyay
Tcet, Mumbai
miloshah55@gmail.com, swatisolanki4206@gmail.com, upadhyay.prakash1999@gmail.com

Abstract – **The variation in solar energy occur daily due to variation in the day-night cycle and also because of seasonal variations throughout the year. The population of the world is increasing very rapidly. From the past decade of years, the non-renewable energy sources like coal and oil are extinguishing and so it becomes a serious problem for providing he reliable energy to the world. But solar energy plays an important source of primary energy. In this project, we propose a dual-axis solar tracking system by which it is possible to catch the maximum amount of solar energy by using Arduino as the main processing unit.**

KEYWORDS: Dual axis solar tracker, Arduino, LDR Sensor, Servo motor.

I. INTRODUCTION

In the present scenario, the variation in the climatic changes has reached a critical level. The main reasons for climatic changes are due to natural causes and man-made destructions like global warming and greenhouse gases are affecting the climatic conditions around the world. In the past decade of years, there is an increase in demand for reliable and abundant electrical energy derived from renewable energy sources renewable energy plays an important role in the energy crisis of the country. The government started to decrease the usage of conventional energy sources and encouraging people to use renewable energy sources like hydro and solar. One such example of renewable energy is solar power. The solar power received by the earth is approximately 1.8*1011MW.The system will tend maximize the amount of power Absorbed by Photo voltaic systems. It has been found that making the use of a Dual axis tracking system, over a fixed system, can increase the power output by 40% - 60%.. In general, the power developed in such applications depends upon the amount of solar energy captured by the collector, and thus the difficulty of developing tracking schemes capable of following the trajectory of the sun throughout the course of the day on a year- round basis has received significant coverage in this project.

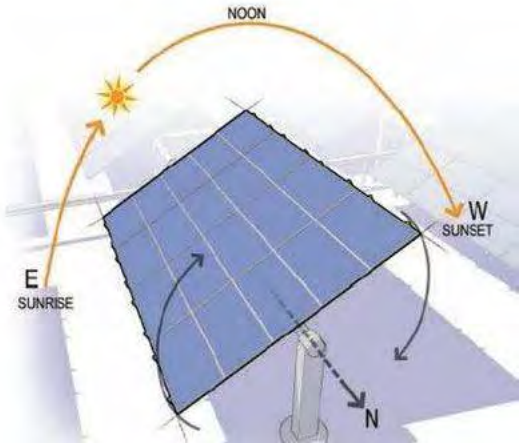
II. SCOPE AND OBJECTIVE

The solar project was implemented using two servo motors.The Arduino IDE was used for the coding. Mumbai has coordinates of 22.5726°N, 88.3639°E and therefore the position of the sun will vary in a significant way during the year. In the tropics, the sun position varies considerably during certain seasons. There is the design of an input stage that facilitates the conversion of light into a voltage by the light-dependent resistors, LDRs. There is a comparison of the two voltages, then the microcontroller uses the difference as the error. The servo motor uses this error to rotate through a corresponding angle for the adjustment of the position of the solar panel until such a time that the voltage outputs in the LDRs are equal. The difference between the voltages of the LDRs is received as analog readings.

The difference is transmitted to the servo motor and it thus moves to ensure the two LDRs are an equal inclination. This means they will be receiving the same amount of light, and the Solar panel will receive the sunlight at 90°, (the plane of PV panel will make an angle 90° with the Sun, and the perpendicular drawn on the plane makes an angle 0° with the Sun, to ensure maximum illumination: Lambert's cosine Law) The procedure is repeated throughout the day.

THE SOLAR TRACKER

A solar tracker is a device which is used to collect the solar energy emitted by the sun. Solar tracking is Nothing but changing the position of panel With respect to the sun. usually, the photovoltaic module assembled in the solar tracker is more powerful than critical irradiance in the fixed system. Solar trackers are classified on basis of performance, coast respectively. by tracking system we can catch 40-50% more efficiency compared to fixed panel. Among them dual axis provides increased efficiency of 48% as compared with single axis tracker



III. HARDWARE REQUIREMENT

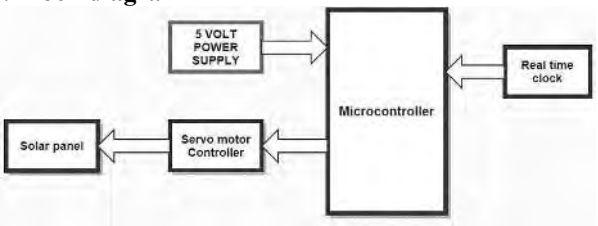
Since it is hard ware based project the main components are LIGHT DEPENDENT RESISTORS(LDR), Servo motors, Arduino as main controller.

A .Light Dependent Resistor (ldr) B .Servo motor C .Solar panel

D .Arduino

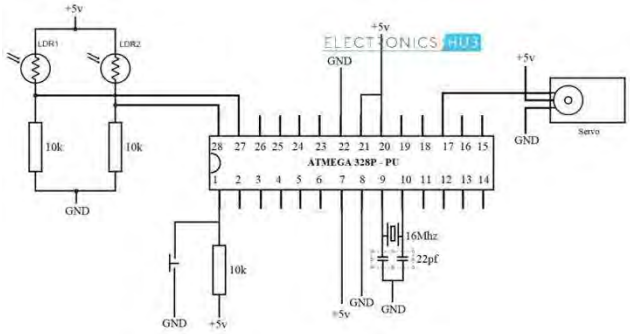
IV. IMPLEMENTATION

A. Block diagram



The principle of the solar tracking system is done by Light Dependant Resistor (LDR). Four LDR's are connected to Arduino analog pin AO to A4 that acts as the input for the system. The built-in Analog-to- Digital Converter will convert the analog value of LDR and convert it into digital. The inputs are from analog value of LDR, Arduino as the controller and the DC motor will be the output. LDR1 and LDR2, LDR3 and LDR4 are taken as pair .If one of the LDR in a pair gets more light intensity than the other, a difference will occur on node voltages sent to the respective Arduino channel to take necessary action. The DC motor will move the solar panel to the position of the high intensity LDR that was in the programming.

B. Circuit diagram



C. Algorithm

Step 1: start the program.

Step 2: initialize all the values.

Step 3: calculate the difference of ldr.

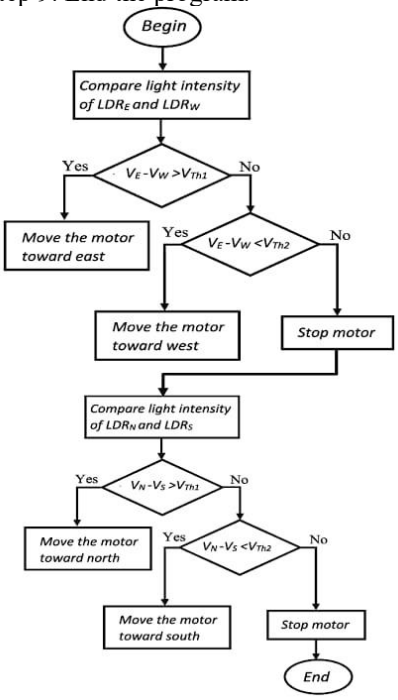
Step 4: if ldr1>>ldr 2,3,4,5servo motor go towards right

Step 5: if ldr2>>ldr 1,3,4,5servo motor go towards centre

Step 6: if ldr1>>ldr2,3,4,5 servo motor go towards right.

Step 7: if ldr1>>ldr2,3,4,5 servo motor go towards right.Step 8: if ldr1>>ldr2,3,4,5 servo motor go towards right.

Step 9: End the program.



Power generated with and with out tracking

Graphical View of Voltage Drawn By Solar Tracking

V. RESULTS

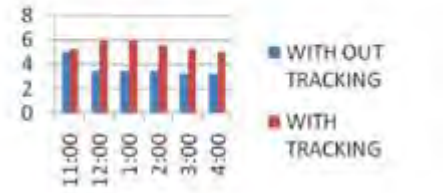
The following tables shows voltage drawn by solar panel with and without tracking respectively.

Time	Voltage without tracking	Voltage with tracking
11:00AM	4.955v	5.262v
12:00PM	3.467v	5.875v
1:00PM	3.447v	6.002v
2:00PM	3.447v	5.515v
3:00PM	3.15v	5.275
4:00PM	3.15v	5.015
Time	Voltage without tracking	Voltage with tracking

11:00AM	3.682watt	3.467watt
12:00PM	4.112watt	2.452watt
1:00PM	4.2watt	2.39watt
2:00PM	3.86watt	2.307watt
3:00PM	3.692watt	2.205watt
4:00PM	3.5watt	2.205watt

Power generated with and with out tracking

Graphical View of Voltage Drawn By Solar Tracking



VI. EXPERIMENTAL SETUP

VII. MERITS AND DEMERITS OF SOLAR ENERGY

Merits:

Once a solar panel is installed, the energy is produced at reduced costs.

Whereas the reserves of oil of the world are estimated to be depleted in future, solar energy will last forever.

It is pollution free.

Solar cells are free of any noise. On the other hand, various machines used for pumping oil or for power generation are noisy.

Once solar cells have been installed and running, minimal maintenance is required. Some solar panels have no moving parts, making them to last even longer with no maintenance.

On average, it is possible to have a high return on investment because of the free energy solar panels produce.

Demerits:

Solar panels can be costly to install resulting in a time lag of many years for savings on energy bills to match initial investments.

Generation of electricity from solar is dependent on the country's exposure to sunlight. That means some countries are slightly disadvantaged.

Solar power stations do not match the power output of conventional power stations of similar size. Furthermore, they may be expensive to build.

Solar power is used for charging large batteries so that solar powered devices can be used in the night. The batteries used can be large and heavy, taking up plenty of spacend needing frequent replacement.

FINALLY,

VIII. IMPROVEMENTS IN RESEARCH

With the available time and resources, the objective of the project was met. The project is able to be implemented on a much larger scale. For future projects, one may consider the use of more efficient sensors, which should also be cost effective and consume little power. This would further enhance efficiency while reducing costs.

If there is the possibility of further reducing the cost of this project, it would help a great deal. This is because whether or not such projects are embraced is dependent on how cheap they can be. Shading has adverse effects on the operation of solar panels. Shading of a single cell will have an effect on the entire panel because the cells are usually connected in series. With shading therefore, the tracking system will not be able to improve efficiency as is required.

XI. CONCLUSION

In this 21st century, as we build up our technology, population & growth, the energy consumption per capita increases exponentially, as well as our energy resources (e.g. fossils fuels) decrease rapidly. So, for sustainable development, we have to think alternative methods (utilization of renewable energy sources) in order to fulfil our energy demand.

In this project, Dual Axis Solar Tracker, we've developed a demo model of solar tracker to track the maximum intensity point of light source so that the voltage given at that point by the solar panel is maximum. After a lot of trial and errors we've successfully completed our project and we are proud to invest some effort for our society

References

[1] Berberi, P., S. Thodhorjani, P. Hoxha, and V. Muda. 2013. Photovoltaics: between a bright outlook and uncertainty. Energy Sci. Eng. 1:72–80.

[2] Chakraborty, S., P. K. Sadhu, and N. Pal. 2015. Paper on Solar tacker.

[3] M.D.Singh and Khanchandani: "PowerElectronics

[4] www. arduino.cc

[5] C. Hua and C. Shen (1998) Comparative study of peak power" trackingtechniques for solar storage system'',Applied Power Electronics Conference

Food Environment Monitoring System

Vinayak Navada, Dhruv Purohit, Dhairya Shah
Information Technology, Thakur College Of Engineering & Technology
vnavada99@gmail.com, purohitdhruv937@gmail.com., shahdhairya1112@gmail.com

Abstract— **The advancement of multipurpose sensors over the final decades has been examined with the point of creating inventive gadgets with applications in a few areas of innovation, counting within the nourishment industry. The integration of such sensors in nourishment bundling innovation has cleared the way for cleverly nourishment bundling. These coordinates' frameworks are competent of giving dependable data around the quality of the pressed items during their capacity period. To achieve this objective, brilliantly packs utilize a assortment of sensors suited for observing the quality and security of food products by recording the advancement of parameters just like the amount of gasses, temperature, stickiness, light exposure and capacity period. This innovation, when combined with IoT, is able to supply a part more information than customary nourishment review advances, which are constrained to weight, volume, colour and perspective assessment. The first framework depicted in this work depends on a basic but successful strategy.**

I. INTRODUCTION

mugginess are significant variables. So the checking framework fit for estimating temperature and stickiness fluctuation during transport and capacity is of prime significance. Today nearly everyone is getting influenced by the nourishment they expend, it's about the shoddy nourishment, yet all the stuffed nourishments, vegetables, items devoured and utilized in day by day life, as every one of them don't offer quality since their temperature, dampness, oxygen content differ now and again. Likewise inappropriate putting away of nourishment materials can make lead wastage of nourishment. Brilliant nourishment observing framework centers around safe stockpiling of nourishment by checking and controlling different parameters influencing nourishment materials.Sanitation and cleanliness is a significant worry so as to forestall the nourishment wastage. The Quality of the nourishment should be checked and it must be kept from spoiling and rotting by the environmental components like temperature, stickiness and dull. In this way, it is valuable to send quality checking gadgets at nourishment stores. These quality observing gadgets keep a watch on the ecological factor that cause or pace up rot of the nourishment. Afterward, the natural elements can be controlled like by refrigeration, vacuum stockpiling and so on.

Fig 1: Block Diagram

Background

II. SUBJECT HISTORICAL DATA

The activity of a control and checking frameworks is to watch out for specific thing or action and to ensure that it remains in the ideal way. Observing can be accomplished utilizing different electronic sensors. Further these recorded qualities can be utilized to control. The information got from sensors can be contrasted with the ideal qualities. In the event that the sensor readings are seen as inconsistent to the ideal qualities, at that point the control circuit will come vigorously to control the allotted action to keep it in wanted way. We recommend utilization of this rule for building a framework which can save crude nourishment. Keen nourishment checking framework is intended to screen and control nourishment materials and keep it from harms happening because of environmental or climatic changes.

Nourishment cleanliness and security is a significant issue for human wellbeing. There are numerous components prompting food contamination, regularly changes in temperature and Despite the fact that we appreciate one of the most secure nourishment supplies inside the planet, foodborne sickness stays regular inside the India where foodborne ailment episodes influence millions and murder a great many individuals. These episodes likewise undermine customer trust in influenced items, end in enormous reviews, and reduce showcase request.

A significant number of foodborne ailment episodes result from poor cleanliness rehearses. Microorganisms can get by in nourishment preparing and dealing with situations. they're by and large brought into the nourishment condition through crude materials, bugs, air, water, and workers. Typically, the normal utilizations of fabulous sanitation rehearses can control these microorganisms inside the nourishment preparing and taking care of conditions. Be that as it may, if tainting levels are high or sanitation methodology are insufficient, microorganisms may build up and _ought to pollute nourishment items prompting foodborne disease flare-up.

Different pathogenic microorganisms like E.coli, Listeria monocytogenes , and Salmonella spp. have settled narratives of being potential contaminants in nourishment taking care of conditions. Consequently, it is imperative to watch the clean condition inside the nourishment fabricating office for the get together of high caliber and safe nourishment items. An ecological observing project (EMP) will evaluate the viability of the general sterile works on during an office and give vital data to forestall conceivable microbial pollution of nourishment items. restrict mind that EMP doesn't make nourishment safe. Or maybe, it gives important information

(source and focus) on marker living beings, waste life forms, and pathogens of worry during an auspicious way.

To diminish the risk of microbial item tainting, one must have understanding into actualizing a proficient EMP during a nourishment office and appropriately deciphering the information to start fitting remedial activities.

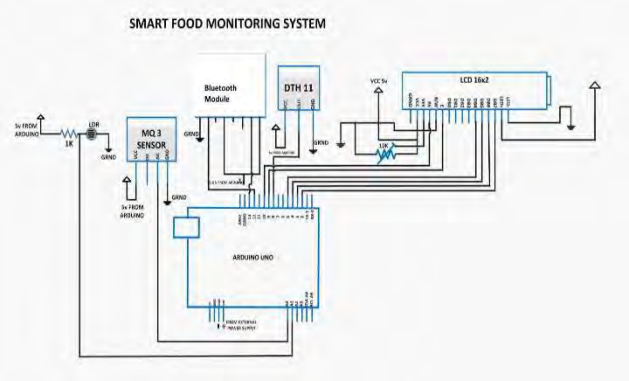


Fig 2: Food Enviromnment Monitoring System Circuit Diagram

Motivation

The various issues related to food incidents occurred in the past year, i.e. 2018. Some of them are as:

1. On 16 July 2013, at least 23 students died, and dozens more fell ill at a primary school in the village of Dharmashati Gandaman in the Saran district of the Indian state of Bihar after eating a Midday Meal kept in environment not favorable for food.

2. In Germany and the Netherlands, meat and milk were found with elevated dioxin concentrations. The dioxin was traced to citrus pulp from Brazil that had been neutralized with lime contaminated with dioxin. 92,000 tons of citrus pulp was discarded. The citrus pulp market collapsed in some European countries. A tolerance level for dioxins in citrus pulp was set by the European Commission. 1999 – In Belgium, animal feed contaminated with dioxins and polychlorinated biphenyls affected more than 2500 poultry and pig farms.

Such incidents created a great loss of life and resources both at the same time Hence, going through the above incidences and motivation to do something for the society we decided to develop a Food Environment Monitoring System using the IoT technology.

III. SUBJECT DESIGN AND CURRICULUM

A. PHASES OF PROJECT

- **Phase 1: Testing and Deployment**

1. **Testing**: System was tested in various environments where the foods are stored like bakeries, cold food storage areas etc and calibrated the readings of the sensors to give the exact output.

2. **Deployment**: The proposed system can be deployed wherever there is a large amount of food is stored and need constant monitoring of the food.

Project planning (Resources, Tools used, etc.):

1. Arduino uno

Fig 3: Arduino Board

The Arduino UNO is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc the board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits.

2. DHT-11 Sensor

Fig 4: DHT-11

The DHT11 is a commonly used Temperature and humidity sensor. The sensor comes with a dedicated NTC to measure temperature and an 8-bit microcontroller to output the values of temperature and humidity as serial data. The sensor is also factory calibrated and hence easy to interface with other microcontrollers.

1. MQ3 Sensor

Fig 5: MQ3 Sensor

The Grove - Gas Sensor (MQ3) module is useful for gas leakage detection (in home and industry). It is suitable for detecting H2, LPG, CH4, CO, Alcohol. Due to its high sensitivity and fast response time, measurements can be taken as soon as possible. The sensitivity of the sensor can be adjusted by using the potentiometer.

Implementation:

IoT device should be installed in a food store. Once it is properly installed and powered on, it connects with the internet via Bluetooth modem and start reading data from the interfaced sensors DHT-11 temperature and humidity sensor, MQ3 and LDR sensor. DHT11 temperature and humidity sensor is a digital sensor with inbuilt capacitive humidity sensor and thermistor it relays a Real- time temperature and humidity reading every 2 seconds. The sensors operators on 3.5 to 5.5 V supply and can read temperature between 0 degree C and 50 degree C and relative humidity between 20% and
95%. The MQ3 sensor senses the presence of gases in the atmosphere and provides the output and thus making us aware about the quality standards of food.

IV. FINAL OUTPUT

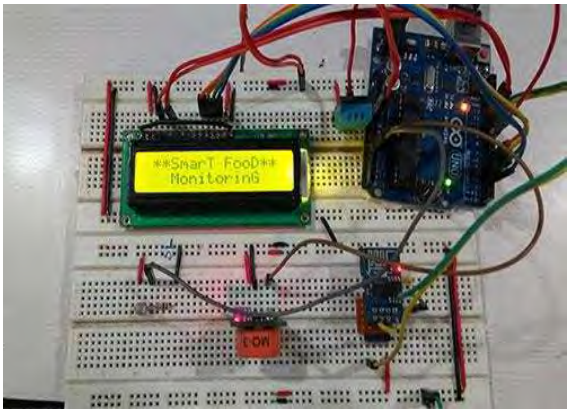


Fig 6: Final Output

The above figure shows the final circuit connection using the block diagram and the circuit diagram. The data is collected using the sensors connected in the circuit and its given to the arduino uno. The arduino uno then compiles the data and the data is displayed on to the serial screen in the computer as well as on the LCD screen

Features of the project

The Project consists of the following features:

1. **Quick Monitoring:** The resulted output is clearly displayed in the LCD (16*2) which is connected in the circuit and hence any fluctuation can be observed through the LCD.

2. High sensitivity and fast response: The alcohol sensor used is highly sensitive to the amount of alcohol generated each point of time in the region as soon as it senses the alcohol above threshold it starts the alert system.

3. **Long life and stable**: The components used are of high grade hence they can be active for a very long time

1.

4. **Simple drive circuit**: The connections and the circuit design of this system is simple hence it is much more convenient to develop

Improvements in Research Gaps (Findings)

Every project has some gaps in it. No project is 100% efficient. There are some gaps in this project too. They are as follows:

1. As this system is currently not WiFi enabled, the data collected is only displayed on the LCD screen. So, respective actions cannot be taken with regard to that. If it is made WiFi enabled, the data can be shown to the owner through a website with a message showing values with c=their respective units.

2. Also, the MQ-3 gas sensor is prone to corrosion due to moisture, which may result in improper functioning of the sensor and can result in serious harm. So, a time-to-time sensor check notification can be sent to the owner, so that a regular check of the sensor is done and improper functioning is prevented.

V. CONCLUSION

In this study, we proposed the food monitoring system that managers and consumers can check the freshness of food. Proposed system consists of temperature sensor, humidity sensor, MQ3 sensor, LDR. To validate the effectiveness of the proposed system, we were monitoring the freshness of milk by the environmental. The freshness of milk can be confirmed through pH at 4°C, 13°C and 22°C. PH of fresh milk is 6.5 to 6.9. And spoiled milk is fewer than 6.5. If improved further this system can be used on almost every store and consumer can get the best quality food and thus improving the life balance. However still lot of improvement is needed in this dynamic real world and we hope that through this prototype there can be further research and future scope possible.

By using this automated system in a cold storage, we can evaluate and monitor the various data by our mobile phone or via PC's. If any abruption takes place then it can be also made automated to turn on/off any external relay module which can operate/condition the temperature and humidity inside a cold storage.

Future Scope

The quantity of IoT gadgets expanded multi year-over-year to
8.4 billion out of 2017 and it is assessed that there will be 30 billion gadgets by 2020. The worldwide market estimation of IoT is anticipated to reach $7.1trillion by 2020. IoT includes broadening web network past standard gadgets, for example, work areas, workstations, cell phones and tablets, to any scope of customarily idiotic or non-web empowered physical gadgets and regular items. Installed with innovation, these gadgets can convey and interface over the web, and they can be remotely observed and controlled

1. Food environment systems can be installed inside the bakery cabins where the food are kept for storage and hence can be continuously monitored and the environment can be maintained for the food kept in them.

2. These food environment monitoring systems can be used inside the cold storages to monitor the environment around the food kept.
3. These monitoring systems can also be used in the restaurants where buffets are used to serve the food.

VI. ACKNOWLEDGMENT

The author would like to thank subject lecturers (current and previous), education designers and other staff members from the School of Computing and Mathematics who previously designed, developed and taught this subject at various times. This work is based on the current author's (who is also the present lecturer and coordinator for the subject) TOL design and development work of the subject.

References

[1]. https://www.mdpi.com/2073-8994/11/3/374/pdf
[2]. http://troindia.in/journal/ijcesr/vol6iss6/73-76.pdf
[3]. http://www.ijircce.com/upload/2017/may/41_A%2 0Survey.pdf
[4]. https://www.oaktreeproducts.com/food monitor- research
[5]. http://medcraveonline.com/MOJFPT/MOJFPT-06- 00194.pdf

An IOT based Smart Medical Box

Divyam Choudhary, Abhishek Gupta, Sanya Gandhi, Shivkumar Chauhan
Department of IT,Thakur College of Engineering and Technology
divyamchoudhary@hotmail.com, abhisekgupta1599@gmail.com, gandhisanya99@gmail.com, cshiv3333@gmail.com

Abstract—**IoT, i.e. Internet of Things, is the currently the most trending field in the Branch of Information Communication Technology. In the medical field when a patient is diagnosed and has been prescribed medication by a doctor, there are chances that the patient forgets to take his medication. When the patient doesn't take the medication on time or forgets to take it, some microbes may survive and they develop resistance to that medicine/drug which leads to antimicrobial resistance (AMR), or drug resistance. The objective of this project is to build a Smart Medicine Box based on IoT technology. It will help in reminding the patients to take their medicines and it will help in monitoring, analysis and controlling the device. Hence we have come up with a smart system that can monitor the patient's medicine intake. The Arduino board will be interfaced with the box that will contain prescribed medicines for the patient. The box will have its own timing information which will be continuously compared to real world time. If the information matches the alarm will start to ring and will remind the patient to take his *medicine.***

Keywords – IOT, Medicine Box, Arduino Uno, ESP8266 Wi-Fi Module, LDR

I. INTRODUCTION

A lot of individuals out in our community need constant help especially elderly people, family members. These are mostly the elderly as they are one of the largest consumers of medicine. But as a human starts to age, different issues arise such as - poor eyesight and poor memory. People may forget to take their medicines on time. In order to eliminate the need for constant observation (through nurses, relatives, guardians, caretaker, etc.) or taking a risk of a missed dose, we had to find an easy, portable and efficient solution. [1]

Nowadays, a modern healthcare IOT platform with an intelligent medical box has a great potential in medical field. Research has found out that about 60% of people above 60 years have poor record of medical history. An intelligent home-based smart medical box will be able to keep the track of medicine intake of the patient and also let the doctor to see the patient's dosage history over the internet. So, if a smart medical box is used which will remind people when to take the medicine, people can keep record of their medicine intake and the doctors will also get a update on daily basis about their patients medicine intake. [2]

The major reason is to bring medical field and technology field closer and make something innovative to reduce the current medical problem. The aim of this study is to build a Smart Medicine Box for reminding the patient to take their medicine at the prescribed time. When the pill time has been set, the medicine box will remind clients or patients to take
pills utilizing light of the LED which will blink when they will set the time to take their medicine.

II. PROBLEM DEFINITION

The Smart Medical Box provides the solution to a very common problem of forgetting to complete a medicinal course by reminding the patients in a very effective manner with proper reminding mechanism.

III. SURVEY

Fig 1: Survey taken with the students at Thakur College of Engineering and Technology. Total 50 students participated

To understand the need of such a reminder tool; a survey of students studying in Thakur College of Engineering and Technology, Kandivali, India was conducted on 19th November, 2019. According to the survey, only 37.5% of students said that they always completed their medicinal course. The survey also emphasizes the statistics that 58.8% had their grandparents on medicinal course who tend to forget to take medicines. The suggestions had obvious solution of a smart reminder.

IV. LITERATURE REVIEW

A Smart Medical Box which can remind us to take the medication has become a necessity nowadays. Before this technology people often tend to forget about taking their medication at the right time which leads to the disease inside the human body get the time to develop antibodies for the medication in that case human would not be able to get treated. It was mainly because patients tend to forget to take their medication in the daily hustle bustle of the life. [3]
Early developments to develop electronic medical boxes started in 2006 by [4]. However, the early electronic medial boxes did not have a monitoring system which can check the intake of the medicines by the patient. Later developments in the electric medical boxes made them smart which uses WIFI modules which connects it to internet and also continuously monitor their status. The most recent advances have added many more functionalities to the smart medical box like it sends the feedback about the pills from the user and send purchase orders to the medical shops.The importance of the Smart Medical Box is starting to become more of a necessity because many patients tends to forget to take their medication on time and with the help of Smart Medical Box which can remind them to take their medicines can really

help them to get cured. [5] Smart Medical Box provides a Stress-Free Reminder
System for taking medicines on time.
The pills are the pillars of medicines, being used as direct treatment of some sort of illness as to assuring some medical condition remains stable.

Failure to take the pills on time will lead to the disease developing immunity to it and patient will not be able to get cured by the same medicine or the same dose of the drug.
We need to carefully make the medical box user face such that it reminds the patient to take the right medicine at right time or sometimes old patients forgets that they have already taken the medicine and they take an overdose of the medicine.

It was found that a large number of medical complications are caused on the grounds that the patients do not take their drugs in the right amount, at the right time. In most cases they either overdose or under dose their medications being forgetful and this deteriorates their health. The patients suffering with Alzheimer have forgetfulness and reduced intellectual capacity as the primary symptoms. This disease is also as a result of expanding age and thus the dominant age group is that of 65 and above. Furthermore, patients may not be able to manage the cost of a caretaker, so dealing with cost viability motivates to build a simple reminder system which is low cost and simple that can be used by a patient [6].

Even patients with less benign illness often forget to take their medicines or sometimes overdose which can cause dysfunction in their bodies. A simple reminder system will reduce the cost but also allow doctors, caretakers, relatives to monitor patients' medical intake. The product can be used for patients who are given home rest through which they can take care of getting their medicines on time and it will not require the help of family members or nurses. It can be upgraded to take the feedback of the medicine from user and the box will be able to place medical orders for the medicine to nearby pharmacies. [4] [7]

Implementation of an IOT based Smart Medical Box which reminds to take medicines on the specified time period and senses the intake of medicines has been done. This Project helps the patients, their families and Doctors to maintain records. The complications caused by incomplete course by patients can now be reduced to a great extent. The circuit has a minimal cost and is affordable in comparison with the manual take care of the patients. This project aims to help all kind of patients with every kind of medicine intake course with proper notifications and reminder. [6]

V. OBJECTIVE

People often forget to take their medicines or sometimes they forget that they have already taken their medicines and they take overdose of the same medicine which can cause dysfunction in their bodies. In order to avoid keeping nurses to check on the patients to take their medicines; it's a user friendly cost effective device which will help the patient to take their medicines on time without fail. It will also allow doctors, caretakers, relatives to monitor patients' medical intake.

VI. REQUIREMENT ANALYSIS

The next step was to understand hardware and software requirements of the project. After looking for a variety of available chipsets and development kits, the Arduino Uno paired with an ESP8266-01 Wi-Fi module were chosen. The cost of the system should also be feasible and hence we came to an understanding for the following components:

Quantity	Components
1	Arduino Uno
1	ESP8266-01 Wi-Fi Module
1	LDR Module
2	LED Lights
14	Jumper Wires
1	Breadboard
1	10K Resistor

Table 1: Hardware Components

As for the software requirement, we use the Arduino IDE which based on the C Programming language.

VII. WORKFLOW AND CIRCUIT DESIGN

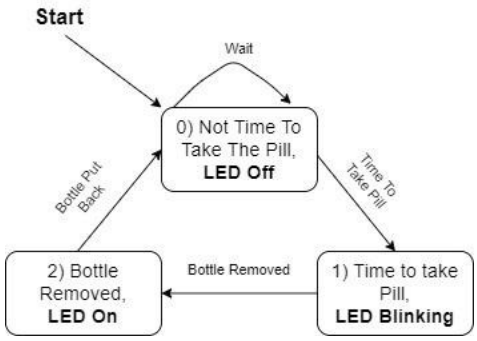


Fig 2: Workflow diagram

The Workflow diagram shows the expected behavior of the IOT solution in a real case scenario. It is depicted in three states:

1. The timer for the medicine reminder is set with the initialization of the system. In this step, the system sits idle and waits for the time of reminder. The LED does not blink.

2. When it is time to take the medicine, the LED will start blinking. This reminds the patient to consume the

medicine. The system continues to stay in this phase until the medicine box is picked up by the user.

3. In this step, the patient picks up the medicine box. The LED stops blinking; the LDR sensor in the system shall senses light from the surroundings and keep the LED turned on. As soon as the medicine is kept back on the medicine box, the LDR cannot sense any incoming light and thus the LED is turned off and the timer is reset again. [8]

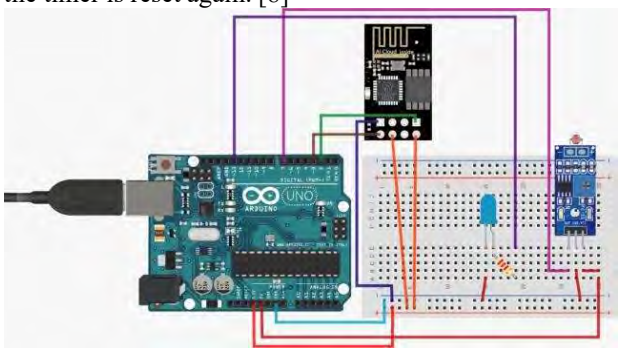


Fig 3: Circuit Diagram

The Circuit Diagram shows how the idea of the workflow is implemented in real time situation. The solution uses Arduino Uno for the circuit. The data collected by Arduino will be light intensity data collected by the LDR sensor. The same data will be transmitted through the ESP8266-01 Wi-Fi module to the Thing Speak API and recorded. Thus, one will be able to record whether the patient took the medicine, or not. Through this data, the patient's family, doctor and patient themselves will be able to see the medicine records and they be relieved of remembering to take their prescriptions.

Fig 5: Smart Medicine Box- LED starts to blink when the time take the medicine was set and it will keep on blinking until the user picks up the medicine on the box.

Fig 6: Smart Medicine Box- As the patient lifts the medicine for intake; the LED will glow continuously until medicine is kept back on place. The system will then again until the next time arrives

VIII. RESULTS

Fig 4: Smart Medicine Box- System is active and waiting for the time of medicine to be taken. LED is turned off.

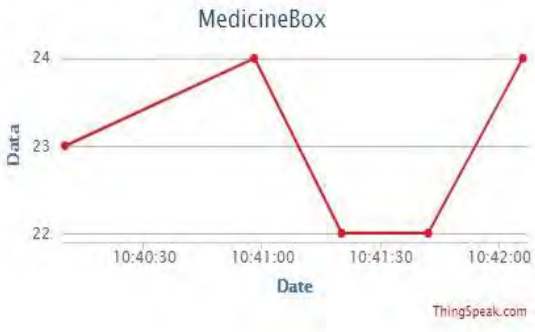


Fig 5: Think Speak Dashboard

As per the purpose of this system, it reminds the patients to intake medicines. After the medicine box is picked up and the medicine is consumed, one can keep the box again at its place; by which there is a cut-off of the light to the LDR module and is a signal of medicine consumption. The recorded light frequency is plotted on the 'Think Speak' Dashboard for the purpose of reporting it to the respective doctor and family member.

IX. RESEARCH GAPS

This being done, there are search some research and technical gaps. Even though this model is working perfectly fine, the user still will not be able to set the time to take its medicine without altering the code. It still needs user input board so that the time can be set by the user manually. As the model is connected with a WIFI module, we can create an android/IOS application for this very model so that the users can control the Smart Medical Box through their mobile phones which will make this model very user friendly. It is also limited to reminding about only one medicine course. A user here has the chances of picking up and keeping it back without the consumption of the medicine.

X. FUTURE SCOPE

Many a times the patient often forgets to take their medication especially old people. Some of them might be suffering from Alzheimer and they always require help from someone else, but what if the medicines they are taking are about to get over and they need new medicines. We can add a device which can detect what kind of medication the patient is taking and the system will automatically detect if the medicines are about to finish or getting near its expiration date and it will automatically send a purchase order to a nearby medical store which will deliver the medicines to the patients' homes.

1. It can have a Mobile Application to make the device more user friendly for non-technical people.
2. Many patients will be able to use it which will make them more self-reliable and they won't be requiring anyone's help.
3. It can be used to send the reports of the patient's intake of medicine to their doctors, and on the basis of those reports the doctor can think of further procedures for the patient.
4. It can be used for more than one type of medicine and remind with a system of adding different medicines with separate buzzers and light indications.
5. A weight sensor can further be added to sense the weight of the medicine box and recheck if the medicine intake is actually done.
6. The device can be linked to online prescription system and will check if all the prescribed medicines are contained in the medicine box.
7. The device can further scan the QR code of the medicine chip notify for expiration of the same.

XI. CONCLUSION

Smart Medical Box is a user-friendly; low-cost which can remind the patients when they should take their medicine was successfully created. Lot of patients who are on a large amount of medications and have a hard time remembering at what time which medication is to be taken will be helped. Doctors can easily keep a record of their intake of through the Think Speak App.

References

[1] F. M. M. A. N. M. N. Hoa Hong Nguyen, ""A review on IoT healthcare monitoring applications and a vision for transforming sensor data into real-time clinical feedback",” in 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD) , 2017.

[2] H. S. ,. Z. P. V. M. Y. S. Sanjay Bhati, ""Smart Medicine Reminder Box",” IJSTE International Journal of Science Technology & Engineering, Volume 3, Issue 10, ISSN:2349-784X, April 2017.

[3] H. C. Y. J. a. G. C. S. Huang, ""The Intelligent Pill Box - Design and Implementation",” in International Conference on Consumer Electronics, Taiwan, 2014.

[4] J. M. H. A. A. a. J. A. K. T. L. Hayes, ""An Electronic Pillbox for Continuous Monitoring of Medication Adherence,",” in International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, 2006.

[5] S. D. Z. Z. a. F. N.-A. A. Sawand, ""Multidisciplinary approaches to achieving efficient and trustworthy eHealth monitoring systems,",” in International Conference on Communications in China (ICCC), Shanghai, 2014.

[6] M. A.-E. Salama Abdul Minaam, ""Smart drugs:Improving healthcare using Smart Pill Box for Medicine Reminder and Monitoring System",” Future Computing and Informatics Journal Volume 3, Issue 2, pp. 443-456, 2018.

[7] G. R. T. Samir V.Zanjala, ""Medicine Reminder and Monitoring System for Secure Health Using IOT”,” in Procedia Computer Science 78 (2016), Nagpur, INDIA, 2015. [8] Chris Forsberg, “Internet of Things in Healthcare,” 29 January 2015. [Online]. Available: https://cforsberg.com/?p=273.

Blood Type Detection Using Photoelectric Sensor And Laser Diode

Suyog Khadke, Ashish Kumar, Rupali Bajpai
TCET , MUMBAI
imsuyog1307@gmail.com, ashishkumar15999@gmail.com, rup.bajpai@gmail.com

Abstract— **The objective of this project is to build a blood type detection device based on IOT technology. The device uses the change in frequency of the light when it is passed through the blood sample to detect the blood type of the patient. Now days the process of blood type detection has become extremely important due to the high blood transfusion rates as well as organ transplantation rates and any mistake in this process can result in the loss of a life that is the patient and due to the above risk the hospitals prefer using only O+ blood type for blood transfusions as they are universal donors but to the ever increasing population using only one type of blood group is not feasible. Thus a device is required which can do the blood type detection process accurately and also with high efficiency. Moreover now a days there has been an increase in wrong blood type determination of a person, this project also take an initiative to overcome that by discussing about it in the future scope**

Keywords–BloodType;Transfusion;Photoelectric Sensor

I. INTRODUCTION

IOT, i.e. Internet of Things currently is one of the most trending fields in the Branch of Information Communication Technology. In case of blood type detection in hospitals and clinics the process which is traditionally used to detect the blood type of a person takes a large amount of time and the precision of that machine is also not very good. The cost of the machine which is
traditionally used to detect the blood type in laboratories is also very high. Therefore for the following reasons there is a requirement of a device which has high precision rate, which is cost effective and provides the result fast.

The objective of this project is to build a blood type detection device based on IOT technology. The device uses the change in frequency of the light when it is passed through the blood sample to detect the blood type of the patient. Now days, due to the ever increasing population the rate of blood transfusions and organ transplantations has increased rapidly and because of this the blood type detection of a person needs to be fast as well as accurate because any mistake in this process can lead to the loss of a life. Therefore for the following reasons a device is required which can detect the blood type in few minutes and that to with high accuracy rate.

After studying the literature survey it was planned to overcome the research gaps in the traditionally used blood type detection devices and the device was tested using the test cases and then the outputs were generated. The output of the device was shown according to different test cases and thus the readings provided by the blood type detection device became more accurate.
A discussion forum is also run for the subject through the subject interact2 page where students can discuss various new and recent topics and can share their experiences with each other. Most students participate actively in the discussion forum and the lecturers just monitor the discussion. Lecturers only participate in the forum discussion when it is required either to provide some answers to a specific question asked or if there is some new information that the lecturer thinks may be useful for a particular discussion.

II. BACKGROUND

Blood type detection has become a crucial aspect in blood transfusions and organ transplantations as opposed to about 90 years ago when automatic blood detection was not developed. Before introduction of blood type detection devices, wrong blood transfusions and organ transplantations has resulted in the loss of human life . Early developments in detection of blood type of a person began in
1901 and the first ABO-System was detected by Landsteiner K in 1901. Further research was done and different subgroups in ABO- System were discovered.

Later developments in the blood type detection systems used the technological advances that have been integrated with the health care industry which in turn increased the accuracy of blood type results thus preventing wrong blood group transfusions and other medical related threats and the time required to get the result was also reduced to great extent.

e subject design and the curriculum. Section IV presents CSU's Graduate Learning Outcomes (GLOs) and their relationship with the subject. Finally, Section V concludes the article.

III. IMPORTANCE OF PROJECT

The requirement of a IOT based device to detect the blood type of a patient is to make the process of blood type detection more fast and also to increase the accuracy of the results obtained.

An average of 200 to 300 blood samples are analyzed within 2 to 3 hours by the traditional blood detection method but when the proposed device is used the

same task can be completed within a short period of time with good precision.

The proposed device can also reduce the human error which can be caused in the laboratories.

The device will cost less than the traditional blood detecting systems hence more cost effective.

You need to carefully consider the benefits that you will get from the device and also the impact of this device in the health care sector.

IV. OBJECTIVES

- The proposed idea is to replace the manual work in clinical laboratories for identifying the blood group.
- To provide a cost effective and affordable method as compared to the traditional method for detecting blood type.

- For large data may misdeed to maintain the blood sample and records accurately. In this new technique, the LASER source and photocell are used to analyse the ABO blood grouping (A, B, O) and Rhesus type testing (Rh Positive and Negative).

It can also be integrated with other sensors to monitor heart pulses and it can also act as a medicine reminder for a patient by integrating a Wi-Fi module and the appropriate sensor for the same.

V. SUMMARY

The proposed idea is to replace the manual work in clinical laboratories for identifying the blood group. For large data may misdeed to maintain the blood sample and records accurately In this new technique, the LASER source and photocell are used to analyse the ABO blood grouping (A, B, O) and Rhesus type testing (Rh Positive and negative)Blood Type Detecting Device ,which helps to save many lives in emergency from Human damage. The project focus on building an IOT based Blood Type Detecting Device System using Arduino as platform. Before the project starts there is the first chapter which speaks about the background of the project, why it is required and also it adds about the need and motivation for the project . This Prototype help the future projects to be very effective and efficient As This is very first time to develop such type of Device in low cost.

04

The Project consists of the following features:

In this new technique, the LASER source and photocell are used to analyse the ABO blood grouping (A, B, O) and Rhesus type testing (Rh Positive and negative).

An average of 200 to 300 blood samples are analysed within 2 to 3 hours by the traditional blood detection method but when the proposed device is used the same task can be completed within a short period of time with good precision.

Blood typing is a test that determines a person's blood type. The test is essential if you need a blood transfusion or are planning to donate blood. Not all blood types are compatible, so it's important to know your blood group. Receiving blood that's incompatible with your blood type could trigger a dangerous immune response.

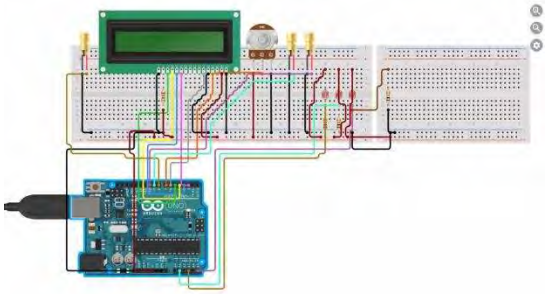
VI. TRADITIONAL APPROACH

The presence or absence of blood stains often provides important information for those investigating criminal cases. For this reason, forensic scientists are often called on to determine whether or not a particular stain is blood, and if so, whose. The detection of blood is usually based on one of three classes of methods. Crystal tests Haem forms crystals when reacted with certain reagents. The most common such reagent is pyridine, which
forms characteristic pink crystals. Catalytic tests these tests rely on the fact that haem can catalyse the breakdown of hydrogen
peroxide. As the H2O2 breaks down, another substance in the reaction mixture is oxidised, producing a colour change. It is important to note that a positive test does not mean that a given stain is blood, let alone that it is human blood, as various enzymes and certain metals can also give positive results. Instrumental methods Chromatography can be used to identify the presence of haemoglobin. These tests are used practically for several different purposes. These include both the
confirmation of the nature of visible stains (i.e. that they probably are or definitely are not blood), the detection of non-visible stains (e.g. on plants or washed clothing) and the enhancement of hard to see stains. Stain enhancement is useful for situations where a footprint, handprint, fingerprint etc. is faintly outlined in blood, as chemical methods can enhance that stain so that the print can be measured and matched with suspects. In all of these tests it is important to ensure that the chemical reactions do not prevent later tests being done to help to identify who the blood belongs to.

VII. IOT BASED APPROACH

Blood sample for which blood group is to be identified is placed in the glass slide between the LASER beam and photocell. After these arrangements, the voltage from the photocell is decreased due to the variations in LASER beam intensity. Then a drop of antigen is added to a drop of blood sample which is on the transparent glass slide. If the clumping (the agglutination reaction) occurs, then the density of the blood sample changes (high) which in turn changes intensity of light to the photocell. If the clumping reaction does not occur, then the density of blood sample does not change. So there will be no changes in the level of energization of photocell According to photocell reading we write program for any microcontroller by using ADC

or comparator to change this voltage in digital value nothing but 0
& 1. Here the circuit diagram of device :



After sensing the input from the sensor it is then send to Arduino and then processed using the following logic table

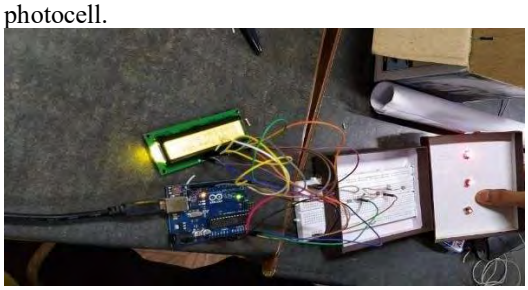
Let us consider a we have sense on first sensor we get the output as high and similarly on other two sensor we get output as low in this case the combine output of case will be A- blood group as it satisfies the all condition .

Anti A	Anti B	Anti D	Blood Group
1	0	1	A +
1	0	0	A -
0	1	1	B +
0	1	0	B -
1	1	1	AB +
1	1	0	AB -
0	0	1	O +
0	0	0	O -

VIII. RESUT & DISCUSSION

Outputs:

1) Blood sample for which blood group is to be identified is placed in the glass slide between the LASER beam and photocell.



2) Lcd display of blood type detection using photocell and laser diode.

IX. OUTCOMES

(1) Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase;
(2) Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done;
(3) Proposing design, architecture for the prototype;
(4) Implementation of the Arduino model with the help code running it in the arduino IDE environment;
(5) Testing was done with the LASER source and photocell are used to analyze the ABO blood grouping (A, B, O) and Rhesus type testing (Rh Positive and negative).The testing of the model was a success are mentioned in the above figures;
(6) Real time sensing was done with the LASER source and photocell are used to analyze the ABO blood grouping (A, B, O) and Rhesus type testing (Rh Positive and- negative)Thus the model works efficiently.

X. FUTURE SCOPE

1) Determining of blood types is very important during emergency situation before administering a blood transfusion. Presently, these tests are performed manually by technicians, which can lead to human errors. Determination of the blood types in a short period of time and without human errors is very much essential**.** The developed method is useful in emergency situation to determine the blood group without human error.
2) The proposed idea is to replace the manual work in clinical laboratories for identifying the blood group.
3) We can design fully automatic machine for this sensor which will help us to improve the response of the system.

XI. CONCLUSION

An average of 200 to 300 blood samples are analysed within 2 to 3 hours by the traditional blood detection method but when the proposed device is used the same task can be completed within a
short period of time with good precision.

One more added advantage of this proposed device is that it is extremely affordable and we can also add extra features such as heart monitoring system etc and thus it can also act as a virtual doctor.(Future project model)

We can design fully automatic machine for this sensor which will help us to improve the response of the system.

References

[1]. Michael J. Hertaeg, Rico F. Tabor, Joseph D. Berry and Gil Garnier, Dynamics of stain growth from sessile droplets on paper, *Journal of Colloid and Interface Science*, 10.1016/j.jcis.2019.01.032, (2019).

[2]. Marek Bialkower, Heather McLiesh, Clare A. Manderson, Rico F. Tabor and Gil Garnier, Rapid paper diagnostic for plasma fibrinogen concentration, *The Analyst*, 10.1039/C9AN00616H, (2019).

[3]. Wang, Ping and Qingjun Liu, *Biomedical Sensors and Measurement*, Berlin Heidelberg: Springer-Verlag, 2011.

[4]. Jones, Deric P., *Biomedical Sensors*, 1st ed. New York: Momentum Press, 2010.

Iot Enabled Smart Parking System

Abhishek Singh, Hrithik Singh
TCET, Mumbai
7227abhishek@gmail.com, hritiksingh1999@gmail.com

Abstract— ***IoT, i.e. Internet of Things is currently the most trending field within the Branch of information Communication Technology. In Residential buildings, malls, commercial enterprise buildings in which parking is an important concern as parking is not easily available. In places with a limited place for parking the vehicles, due to ineffective management, parking becomes an issue the objective is to build a smart parking system which will allow effective and consistent solution for the parking issue. Nowadays Smart parking is a demand for the management of these places because of the populace which is increasing with the aid of day. Also, as of now with the dropping expenses of automobiles parking is a real hassle. due to ineffective parking, it leads to visitors jam which prices everyone their time. So, with powerful management, this hassle can be reduced quite a bit. After studying the literature survey, it turned into a plan to triumph over the studies gaps and the system become tested the use of a model with dummy cars. A smart parking system can help to resolve this trouble. In a smart parking system while the parking spot is empty it's going to let the car in and if any parking spot isn't empty it's going to let the driver recognize the parking spot at the entrance through the LCD screen. Moreover, there may be an increase in these scenarios nowadays, this mission additionally takes an initiative to conquer that by discussing it in the future step.***

Keywords – Arduino UNO, IR Sensor, Servo Motor, Battery, Connectimg Wires, Automation, Smart parking etc

I. INTRODUCTION

Internet of Things or IoT is a widely - used enterprise buzzword offering widespread transformation possibilities for corporations. An IoT solution includes sensors and actuators, communication channels, statistics collection at the cloud, and applications. It comes with its demanding situations of scale, diversity, connectivity, privacy, safety, and regulatory compliance. IoT platforms enforce common reference architecture comprising the Edge, Platform and Enterprise tiers and help groups triumph over technical challenges. Any smart applications include sensors, which are deployed in the environment, collect the information from the device/sensor is processed and analyzed to manage the applications. This approach would reduce the cost of manpower and increase productivity. Commercial buildings, residential homes, etc. Nowadays are much more crowded, due to this population increases the want for powerful management of parking machine. A smart parking machine can help clear up this problem. In clever parking gadgets when the parking spot is empty it'll permit the car to enter and if any parking spot is not empty it will permit the driver to know about the parking spot at the entrance through the LCD screen. Moreover, there may be a boom in these eventualities nowadays, this undertaking additionally takes an initiative to overcome that with the aid of discussing it in the future step.

A. Background

The automated car parking system refers to a mechanical system devised to park a largen number of cars with maximum efficiency. Automated Parking System was first developed in 1905 in Paris, France at the Garage Rue de Ponthieu. The necessity for an automated parking system works based on one's need for a space to park and the undersupply of it. With our focus firmly on customer experience, our Australasian-based technology division, alongside a rapidly growing global network of expert partners, aims to bring our intelligent parking system and integrated smart parking solutions to the world. this example might be taken as a chance so as to make advancement within the efficiency of parking resources which will reduce the searching time, traffic jam and road accidents The System is mainly used in Europe, the USA and Japan (Shaheen et al., 2005) is evolved with the incorporation of advanced technologies and researches from various academic disciplines. With its deployment inside the car park, it's far hoped that it'd solve the aforementioned troubles faced by the patrons within the car park.

B. Importance of Project

The requirement of the smart parking system is because if the increasing population of the cars at the commercial building, residential building's mall, etc. Hence there is great importance for this project in Day-to-Day life. Importance of the smart parking system are as below

- Use new technology that uses Internet of Things (IOT).
- LCD screen will let the driver know about the empty spot so they don't waste time in finding the parking spot.
- Effective management will not make the area crowded.
- As soon as spot is emptied it change the status in the LCD screen which will inform the driver about the status.
- To achieve cost savings.
- Human force would decrease because of automatic opening and closing of the gate

C. Motivation

The increase in town traffic is one of the foremost effects of growth, particularly in urban areas. because of this looking for a vacant car park throughout peak hours isn't solely long however additionally leads to wastage of fuel. A Smart Parking System like this enables drivers to construct good choices that may cut back congestion and construct the foremost of accessible areas. Finding a car parking zone has come to be a daily challenge of late, which is anywhere the incentive for this assignment came up. With the evolution of technology, we've smartphones, sensors that find the presence of any object and my plan has a system anyplace parking regions are ready with the Infrared sensors that tell

concerning the occupancy standing of the parking spaces on the LCD screen this is displayed out of doors the park. The drivers keep finding out appropriate automobile parking space that increases traffic. Increasing the volume of transport exhaust creates a negative impact on the surroundings. hence a smart parking system has become the requirement of the day

D. Objective And Scope

The Objective of this project is to implement a low cost, reliable parking system which helps in efficient parking and utilization of space it also helps to scale back pollution and traffic. The Scope are often an android based application are going to be used with a connected database through which a user can book a parking slot consistent with his time.

- To reduce with continuous effective flow of parking spot.
- To provide safe and secure parking slots within limited area.
- To scale back pollution by decreasing drive time.
- To save time by making parking easily available.

II. LITERATURE SURVEY AND PROPOSED WORK

In this section, we present literature survey done related to this project and relating work in progress.

A. Introduction

Various methods are prevalent for the development of autonomous or intelligent parking systems. The study of these systems shows that these require a little or more human intervention for the functioning. One of the intelligent systems for car parking has been proposed by making use of IR sensor LCD and servo motor. In this work, I have developed a Smart Parking System that can monitor available empty slots economical and reliable and consider contributing to fuel and time-consuming. In this work, the most widely used sensors are Infra-Red (IR) sensors.[3] These sensors are used for monitoring the vacancy of a parking slot. The reason for using these sensors is they are cheap and use less memory as compared to the camera if used as sensors. The sensor will monitor whether the car is parked or the slot is empty. If the slot is empty, then the sensor would sense it and give the notice to LCD so in the LCD the vacant parking slot would be displayed Servo motor is used as gate for opening and closing of the gate tail Thus, we aim to propose a car parking system that represents a fully automated model with minimum human intervention and overcome the limitations of existing system

B. Literature Survey

From this research paper we got definite data related to problem caused by traffic. A Reservation – based Smart Parking System by Hongwei Wang and Wenbo He – April 01, 2016. We got information about how IOT based smart parking can be implemented with help of sensors and how cloud storage can be used for Y. Rahayu, F. N. Mustapa, ":A secure parking reservation system using gsm technology a smart and secure parking reservation system based on GSM technology.

C. Problem Definition

To design an IOT based SMART PARKING SYSTEM that will continuously monitor the presence of the car at the parking and show the appropriate output in the LCD at the entrance. Smart parking may be a solution to metropolitan cities to scale back congestion, cut vehicle emission totals and save persons' time by helping them find a spot to park. Smart Parking may is and advanced parking system, usually a parking system replacement of the traditional parking system. Smart parking system equipped with special sensors to detect the available parking slots at any parking lot.This is an application supported Internet of Things (IoT) that in Real-Time environment have sensors and devices embedded into parking spaces, transmitting data on the occupancy status; and the vehicle drivers can search for parking availability using the information displayed on the LCD screen. Hence the driving force would know where there's an available spot to park his vehicle in less time, reducing the energy consumption and pollution

D. Phases of Project

- Phase 1: Arranging, Examination, Planning and Execution.

Arranging: Arranging: After examination, we learned about our subject and did some exploration on it for our better comprehension of the undertaking and furthermore to get an unpleasant picture about what might be our concern definition for the specific venture.

Examination: We got the away from our task title by doing research on it and will conclude our definition and after that, we previously made the Writing Overview of the venture and did the entire documentation.

Structuring: At that point, we will develop the plan of the task and as per that, will list down all the necessities required for the development of the model of our venture.

Execution: In the wake of obtaining the necessities, we continued for the development of the application.

- Phase 2: Testing and Arrangement.

Testing: After the model is prepared, we will initially interface the equipment with the allocated code and afterward we will check in the event that it underpins the instrument or not. If not, we will illuminate the issues with respect to it and will check once more.

Arrangement: After, complete coordination and testing of task ongoing running and activity of the framework will be finished. We are relied upon to approve against their recently put away information.

III. ANALYSIS AND PLANNING

Getting a clear plan of the project title and doing analysis thereon we are going to get our definition and at that time then we are going to 1st produce the Literature Survey of the project and do the total documentation. once analysis we are going to 1st study concerning it and do some analysis

thereon for our higher understanding of the project and conjointly obtaining a rough image concerning what would be our drawback definition for the actual project.

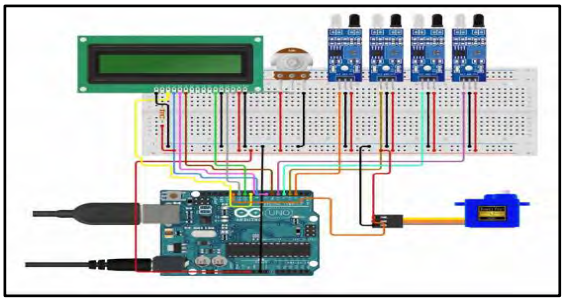


Fig.01 Circuit Diagram

A. Feasibility Study

1. Technical Feasibility: Here one has got to check, whether or not the projected system is developed victimization existing technology or not. It's evident that the mandatory hardware and software package area unit obtainable for the development and implementation of the projected system. Hence, the answer is technically possible.

2. Economic Feasibility: As a part of this, prices| the prices} and advantages related to the projected system compared and therefore the project is economically possible given that tangible or intangible advantages to outweigh costs. System development prices are going to be important. So, the projected system is economically possible.

3. Legal Feasibility: Legal problems will affect a system's acceptance by users, its performance, or the selections on whether or not to use it within the initial place—so it's best to contemplate these expressly in system style. The behavior of these being registered and recognized will influence the accuracy and effectiveness of just about any biometric system.

4. Operational Feasibility: it's a regular that ensures ability while not stifling competition and innovation among users, to the advantage of the general public each in terms of price and repair quality. The projected system is suitable for users. So, the projected system is operationally possible.

B. Implementaiom Details

Arduino Uno (Fig.02): It is a microcontroller board dependent on an 8-piece ATmega328P microcontroller. Alongside ATmega328P, it comprises of different segments, for example, precious stone oscillator, sequential correspondence, voltage controller, and so forth to help the microcontroller. Arduino Uno has 14 advanced information/yield pins (out of which 6 can be utilized as PWM yields), 6 simple information sticks, a USB association, A Force barrel jack, an ICSP header, and a reset button.

Fig.02 Arduino Uno

Fig.03 IR Sensor

IR Sensor (Fig.03): As referenced that the model is IR sensor situated it deals with some standard. Infrared sensors deal with the guideline of reflected light waves. Infrared light reflected from articles or sent from an infrared remote or signal. Infrared sensors are additionally used to quantify separation or closeness. The reflected light is recognized and afterward a gauge of separation is determined among sensor and article.

Fig.04 Servo Motor

Servo Motor (Fig.04): The servo motor position can be controlled more correctly than those of average DC engines, and for the most part, they have three wires like force, GND, and control. Capacity to these motors ceaselessly applied, with the servo engine control circuit changing the attract to drive the servo engine. These engines are intended for increasingly careful assignments where a motor position should be clear absolutely like moving a mechanical arm or controlling the rudder on a pontoon or robot leg inside a specific range.

Fig.05 LCD Screen

Fig.06 Potentiometer

LCD screen (16*2) (Fig.05): LCD modules area unit normally utilized in most embedded comes, the rationale being its low-cost value, availability, and computer programmer-friendly. Most people would have bumped into these displays in our day to day life, either at PCO's or thus because; it's 16 Columns and a couple of Rows. There are unit tons of mixtures out there like, 8×1, 8×2, 10×2, 16×1, etc. however the foremost used one is that the 16×2 alphanumeric display. So, it'll have (16×2=32) thirty-two characters in total and every character will be manufactured from 5×8 component Dots.

Potentiometer (Fig.06): A potentiometer could be a three-terminal electrical device with a slippy or rotating contact that forms Associate in Nursing adjustable potential divider. If solely 2 terminals area unit used, one finish and also the wiper, it acts as resistance or resistance. The measuring device referred to as a potentiometer is a potential divider used for mensuration voltage (voltage); the part is Associate in Nursing implementation of an equivalent principle, hence its name.

C. Features of the project

The project contains the following features:

1. It ensures fast and automatic parking and simple retrieval of vehicles.

2. best suited for parking in offices, malls, and similar places.

3. Low maintenance levels area unit needed by the system.

4. Sensors used have high sensitivity and area unit simple to handle

5. The low price system, providing the most ` automation.

6. It doesn't need line-of-sight operation.

7. Friendly reorientation of cars for driving in and out.

8. Safety of automobiles.

D. Methodology used Algorithm

Step 1: Start

Step 2: Initialize IR Sensor, Servo Motor and LCD screen.

Step 3: If car approaches to the gate IR detect and the gate opens

Else it remains close.

Step 4: If: car is parked in the slots show the slots full and increase the count by one.

Else: car is making an exit decrease the count by one.

Step 5: Stop

IV. RESULTS AND DISCUSSION

A. Output

Based on the sensor use, we are able to achieve the desirable outputs as follows:

Based on the sensor use, we can achieve the desired outputs as follows:

1. When the infrared sensor detects the approaching vehicle it sends the response Arduino board

2. As the board detects the info from the Arduino the servo motor acts accordingly.

3. As the vehicle enters the parking arena the total count of vehicle increases by one which is displayed on the LCD screen.

B. Outcome

Phase one Analysis: Creating the outline, literature survey associates degreed practicableness study for the planned resolution which can act as an input to the planning section.

Phase one Planning: Proposing Statement of labor, scope definition and scope boundary for definition and scope boundary for designing the paradigm from the matter definition to come to a decision what has to be done and what to not be done.

Phase one Designing: Proposing style design of Arduino and management mistreatment IoT system for implementing the surveyed case studies and outline.

Phase one Implementation: Various modules are created like the affiliation of sensors. Completion of cryptography grouping information from sensors.

Phase 2 Testing:

Case 1: IR obstacle sensor connection

Components/Things Required: IR Sensor, Arduino UNO.

After connections as shown in the circuit diagram, whenever the car was detected by the infrared sensor it gave output to the Arduino.

The IR sensor was tested successfully and was in working condition.

Case 2: Servo motor.

Components required: Servo motor

After connection as shown in the circuit diagram, the predicted results were successfully observed. Whenever servo motor received input from Arduino it worked correctly and raised its arm.

Case 3: IR sensor.

Components required: IR obstacle sensor

After connection as shown in the circuit diagram, the predicted were tested successfully observed. Whenever the car was detected through these sensors the LCD screen showed the appropriate parking spot.

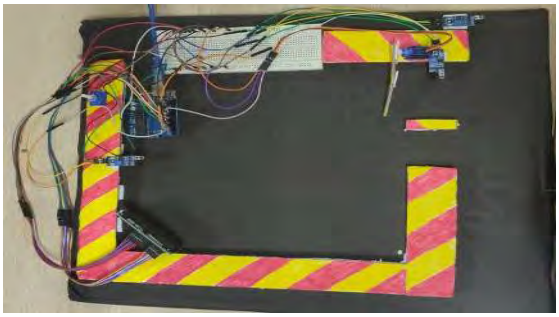


Fig.07 System Design

V. CONCLUSION

The system provides a true-time method and data of the parking slots. This paper enhances the performance of saving users time to find AN acceptable parking lot Analysis of the model needs to be done whereas developing a life-size model. The mechanical model has been designed and also the package still because the negative feedback circuit has been enforced with success. It demonstrates the operating of the planned machine-driven sensible parking system. the most benefits are house improvement, value effectiveness, and security. The code was enforced exploitation Arduino IDE and also the code was dead with none trouble. The circuit diagram was designed exploitation Circuito.io and by referring the circuit diagram the connections were integrated adequately. As for the long run work the users will book a parking lot from an overseas location. GPS, reservation facilities and vehicle plate scanner will be enclosed within the future.

References

[1]. J. DongjiuGeng, Yue Suo, Yu Chen, Jun Wen, Yongqing Lu, Remote Access and system supported Android Mobil Phone, vol.2. Journal of Computer Applications, 2011, pp. 560-562

[2]. J. DongjiuGeng, Yue Suo, Yu Chen, Jun Wen, Yongqing Lu, Remote Access and system supported Android Mobil Phone, vol.2. Journal of Computer Applications, 2011, pp. 560-562

[3]. Y. Rahayu, F. N. Mustapa, ": A secure parking reservation system using gsm technology a smart and secure parking reservation system based on GSM technology.

[4]. Available online through the following url: https://www.aurdino.cc

[5]. Available online through the following url: https:/electronicsforu.com

Dynamic traffic light system

Kashish Jain, Khem Agarwal, Rohan Singh
Thakur College of Engineering and Technology, Mumbai India
jainkashish37@gmail.com, khemagarwal1@gmail.com, rohanrsingh9@gmail.com

Abstract— **Our project presents a traffic controller that m anages traffic flow in a group of intersections in an autono mous, centralized and optimal manner... The device obtai ns information from a network of cameras, and can detect the number of vehicles in each of the roads by machine vi sion algorithms. The Fuzzy system uses this information to select the sequence of phases that optimize global traffic fl ow. To assess controller performance, a scenario was developed where two adjacent intersections could be simulated through artificially created videos. Comparison of system performance vs. fixed time controller. The total waiting time of each vehicle was used as control variable. The results show that the performance of the sys tem increases by about 20 percent over situations with hea vy traffic conditions and that the controller can adapt smo othly to different flux changes.**

Vehicle traffic is on a tremendous increase around the globe, especially in large urban areas. In this study, inspired by recent advanced vehicle technologies, we take account of the improvement of the flow of traffic in real- time problems. To solve the problem, we propose a new approach for managing traffic flow at the intersection in real time via traffic light scheduling control. The proposed approach is based in particular on the theory of process synchronization and connected vehicle technology, in which each vehicle can interact with others. In case of high traffic volume the traffic deadlock is also taken into account. The simulation shows the possible results compared with the current traffic management system.

Keywords: Object Detection, traffic, signal, traffic lights.

I: INTRODUCTION

Nowadays world possesses a serious mobility problem, which affects a great part of the citizens and harms drastically its productivity and competitiveness. One of the main reasons, which contributes to this situation, is the use of inefficient and obsolete traffic controllers, which are not capable to manage in an efficient way. These fixed time controllers, require a periodical configuration based on statistical flow analyses, which generally do not reflect in an accurate way the real traffic flow conditions. To solve the problem we have come up with many solutions. Although the performance of these systems easily exceeds the performance of fixed time controllers, they present a maintenance problem mainly concerning the kind of sensors used. This kind of deployment leaves them exposed to all kind of physical interactions, which reduce drastically their useful life. To avoid this problem, our system was developed, which is capable of managing the flow of the traffic in a controlled manner.

II: BACKGROUND

Below, the most relevant terms and investigations in the area of traffic controllers and vehicle detection inside images are mentioned.

A. Traffic controllers there are two main kinds of traffic controllers: static ones and dynamical ones. The first ones are those where a sequence of actions previously programed are followed, while the other kind makes use of a certain acquisition method, which allows the system to identify the state of the traffic flow on the roads and guide his actions to optimize the traffic flow. On the other hand, it is important to define a basic terminology: phase, cycle and coordination. A phase is a traffic signal which allows a flow of non-conflictive movements. For example, Phase 1 showed in Figure 1 allows traffic flow from west to east and vice versa. In the same way, a succession of phases which is repeated continuously is considered a cycle. Finally, coordination is the action of programing the signalized intersections in such a way, that the flow of a corridor can achieve a constant speed without detentions, generating what is known as green waves.

B. In object detection field, there are two main strategies concerning the vehicle detection task: the first one is based on background and optical flow estimation, while the second one uses machine learning techniques. Background estimation analyzed the difference between a predefined model (image) of an empty road and an image of the incoming traffic, obtaining perturbations that overlapped to the predefined modern interpreted as vehicles. Examples of methods used within this area are: Boosted Cascade of hear Features, Sift (Scale Invariant Feature Transform) matching and neural networks Portion of the work is used in helping the scenario which will be very helpful for further evaluation. Classifier is an operator which uses the features of a data set, identifying the class or group to which each of these data belongs. Algorithms are used for boosting the content and accuracy if the system which will be used for further evaluation. There are plenty of investigations in the area of vehicle detection through images; the following are some of the most important researches in this field: an on road vehicle detector was developed through a Hear feature detector, obtaining an accuracy detection of 88, 6% and 76% respectively.

III: IMPORTANCE OF THE PROJECT

The requirement of the dynamic traffic light system is because of the increase in the pollution due to traffic on the traffic junctions. Hence, there is great importance of the project in day-to-day life. Importance of the Dynamic traffic light system are as below:

Dynamic traffic light system are primarily designed to reduce pollution caused due to static cars present on the traffic junction.

Correctly maintained and operating dynamic traffic light system are effective and proven lifesaving.

Dynamic traffic light system also saves time of all the passengers as the traffic time reduces.

Air pollution is reduced drastically.

Even noise pollution is being reduced.

An effective flow of traffic is crated.

Seamless traffic is being achieved.

IV: MOTIVATION

As the population of the world increases day by day the number of vehicles on the road also increases. As the number of vehicles increases more and more the traffic also increases. Which in turn increases the pollution in the environment.

So our main motive towards this project was a social cause that is to reduce pollution in the environment. So by reducing the pollution we are protecting our mother earth but also matter of ease and benefit to the people to have a smooth and faster commute.

V: OBJECTIVES AND SCOPE OF THE PROJECT

Our main objective is to create a dynamic traffic light system so that it can reduce pollution by reducing traffic on roads. With the help of cameras attached to every traffic junction we will apply object detection algorithms to each and every lane and will give appropriate timers to each and every lane.

Vehicular traffic keeps on increasing day by day. This have been a real issue in this world. Improving the traffic flow has become a necessary gesture towards the society.

Each of the above-mentioned topics has its own web page in the subject interact2 site. The format of the topic is as follows. It starts with the learning objectives, topic overview, suggested readings and activities that may be some hands-on projects followed by a quick quiz to test the understanding, discussion topics for in the class and for the online forum (through subject interact2 site) and then some useful web links providing latest

References related to the topic. All topics follow a similar pattern.

VI: SUMMARY

Dynamic traffic light system are primarily designed to reduce pollution caused due to static cars present on the traffic junction.

Correctly maintained and operating dynamic traffic light system are effective and proven lifesaving. So our main motive towards this project was a social cause that is to reduce pollution in the environment.

So by reducing the pollution we are protecting our mother earth but also matter of ease and benefit to the people to have a smooth and faster commute.

Not only reducing the pollution but also reducing the wait time of the people is our concern.

The requirement of the dynamic traffic light system is because of the increase in the pollution due to traffic on the traffic junctions

VII: FEATURES OF THE PROJECT

The Project consists of the subsequent features:

1. Led lights: We are providing display of traffic lights with the assistance of led bulbs which can indicate proper traffic lights.
2. Fast response: the thing detection model used may be a high accuracy model which provides a faster response.
3. Simple drive circuit: The connections and therefore the circuit design of this method is straightforward hence it's far more convenient to develop.
4. Detection of traffic is carried out.
5. Object detection models are used to detect the traffic density.
6. The type of vehicles are also detected in our system.
7. The system will be more efficient with many factors being considered into the system

VIII: METHODOLOGY

The steps of the execution of our project or the system are being considered in according to the steps.

These steps are mentioned below:

1. Firstly, with the assistance of cameras attached to the traffic signal junction we are fetching the photographs then processing it.
2. Once the processing is completed we get the count of the vehicles. After the count we set the timers for every and each signal supported the count
3. If the count of a sign is high we'll set the timer more and if it's low we'll keep the timer low.
4. We'll change the traffic controller in an exceedingly cyclic manner in order that no lane lefts out.
5. Traffic flow is a real issue. We have developed a system which will detect it and handle the request.

6. Traffic efficiency is a big issue which can be tackled by many means.
7. Main means are only two which are traffic flow and timer.
8. Only making traffic timer is not sufficient.
9. We have to develop traffic flow accordingly to accommodate traffic density.
10. Once the processing is completed we get the count of the vehicles.
11. After the count we set the timers for every and each signal supported the count.
12. If the count of a sign is high we'll set the timer more and if it's low we'll keep the timer low.
13. If there are many cars on one side of the lane we will give that lane preferences.
14. Other lanes will not be given preference.

IX: BLOCK DIAGRAM

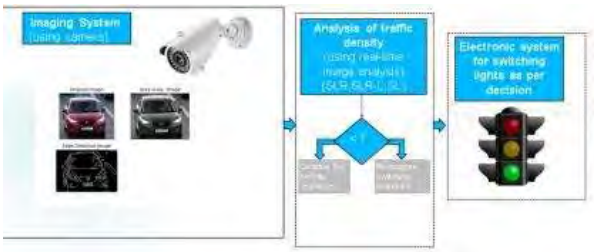


Figure 1: Block Diagram

X: SCREENSHOT

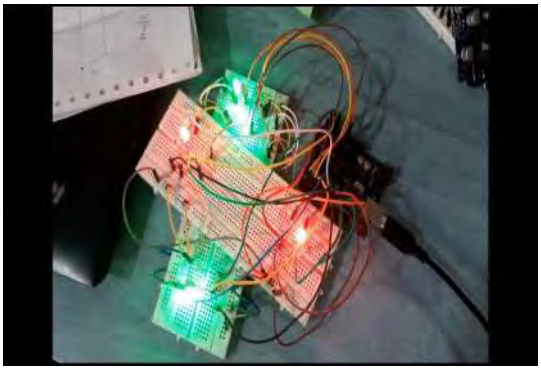


Figure 2: Screenshot

XI: DETECTING CAR USING OBJECT DETECTION:

Figure 3: Object Detection

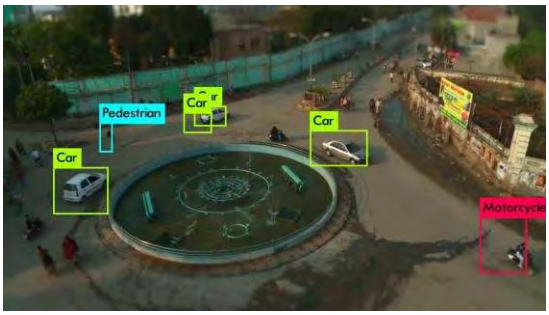


Figure 4

XII: CONCLUSION

We have successfully demonstrated vehicle detection by using object detection algorithms that can detect the vehicles on every traffic junction lane.
Once the analysis is complete we get the vehicles count. After count we set the timers based on count for each and every signal. If a signal count is high we set the timer higher and if it is low we keep the timer down.
We'll cyclically change the traffic controller so no lane is left out. So we completed successfully the IOT-based project model.

References

[1]. Implementacion American state UN sistema American state simulacion para el management Del trafico automotor.
[2]. Distributed and Cooperative Fuzzy Controllers for Traffic Intersections cluster.
[3]. IEEE Transactions on Systems, Man, and information processing, Part: C. Vol.
[4]. Management y simulación American state tráfico urbano linear unit Colombia: Estado del arte.
[5]. Revista American state ingeniería #29, Universidad American state los Andes.
[6]. Speedy Object Detection employing a Boosted Cascade of straightforward options.
[7]. Associate in Nursing Extended Set of Haarlike options for speedy Object Detection.
[8]. Empirical Analysis of Detection Cascades of Boosted Classifiers for speedy Object Detection.
[9]. WIERING, Marco, VAN VEENEN, Jelle, VREEKEN, Jilles, KOOPMAN, Arne.

Smart Agriculture

Aditya Pai, Siddham Shah, Ritik Bohara
Thakur College *of* Engineering *and* Technology
aditya.bondgai@gmail.com,sidshah628@gmail.com,ritik2701.rb@gmail.com

***Abstract*— Internet of Things (IoT) technology has brought revolution to each and every field of make a self-configuring network. The development of Intelligent Smart Farming IoT based devices is day by day turning the face of agriculture production by not only enhancing it but also making it cost-effective and reducing wastage. The aim / objective of this paper is to propose a Novel Smart IoT based Agriculture Stick assisting farmers in getting Live Data (Temperature, Soil Moisture) for efficient environment monitoring which will enable them do smart farming and increase their overall yield and quality of products. The Agriculture stick being proposed via this paper is integrated with Arduino Technology mixed with various sensors and live data feed can be obtained. The product being proposed is tested on Live Agriculture Fields giving high accuracy over 98% in data feeds.**

***Keywords* – Agriculture IoT, Agriculture Precision, Arduino Mega 2560, DHT22 Temperature Sensor, Smart Farming, Soil Moisture Sensor,**

I. INTRODUCTION

Today, individuals have access to an array of smart applications like smartphones, laptops, smart televisions, refrigerators, microwave ovens, etc. However, there aren't many smart devices for the bedroom department of the house. One prominent and important feature that exists in the bedroom is a wardrobe or closet. In fact, over 80% of the respondents we surveyed agreed to have access to a wardrobe. Smart wardrobe helps users manage their clothes inside their wardrobe. Besides having an attached screen to show the status of each cloth, it can also push all data into a cloud and further sync up data with a mobile application. User can get suggestions on what to wear today from their phone app based on the events in their calendar and the weather.

India is a country where majority of our population are dependable on the agriculture to live their daily life. In this modern technological era poor farmers of india cannot get enough assistance from others to help them with technology and make their work easier. This project made automatic field monitoring & controlling system that can be utilize to improve the condition of green houses. Arduino Uno microcontroller is the main controlling unit of whole unit

our goal is to put together a set of technologies into a system that could be used to aid the farmers to locate the defect in their farms

PROPOSED SYSTEM

To design a microcontroller (IoT) based Smart Agriculture that will detect the soil moisture and tempreture and thereby, will help farmers to locate the problem in their fields easily.

II. BACKGROUND

The demand for smart technologies such as Big Data, cloud-based services, GPS, and the IoT is gaining pace in the agriculture industry. Driven by the rising need for high precision crop analysis, automated farming techniques, and collection of data from the field, the world is likely to witness the agriculture industry get smarter with the implementation of aforementioned technologies in the coming years. Data thus derived from implementing smart technologies can help farmers yield high quality and larger quantity of crops.

The rapid escalation of food demand due to the growing population worldwide is boosting the demand for smart agriculture. The smart agriculture engages advanced technologies such as Big Data, GPS, IoT and connected devices. Smart agriculture helps in automated farming, collection of data from the field and then analyses it so that the farmer can make accurate decision in order to grow high quality crop. The field data are collected with the help of sensors, cameras, micro controllers, and actuators. Then the collected data are transferred via internet to the operator or the farmer for decision making.

III. IMPORTANCE

To survive in the highly competitive and rapidly changing agriculture market of the twenty-first century, a farmer's passion for working on the land and hard work is no longer enough. Innovative farming technologies require more and more professional skills.A modern farmer should be an expert in agricultural regulations, data analysis, accounting, budgeting and now in technology too.The good news is that as more new farming technologies emerge, they get more affordable. There are many free apps for farmers who are new to tech. What's more, most of the smart farming products offer exhaustive educational material since thoughtful tech providers adapt their products to the needs and capabilities of farmers.

IV. OBJECTIVES AND SCOPE OF THE PROJECT:

Soil Moisture and Temperature detecting machine

The main objective of this project was to design a greenhouse monitoring system that is be highly reliable and is useful for harvesting crops. Our project mainly focuses on the control of parameters such as-temperature, methane quantity, soil moisture Can fit perfectly in every house and in daily life of people

1. Supply water according to moisture level of soil.

2. Automatic alarming system to avoid the burning of plants by excessive temperature of atmosphere

3. Reusing process of excessive water in the field..

V. FEATURES OF THE PROJECT

The Project consists of the following features:

4. Simple drive circuit: The connections and the circuit design of this system is simple hence it is much more convenient to develop.

5. Time saving: the project helps us to save time by doing all the search work with the help DHT22 Tempreture and Moisture sensor.

3. Hassle-free: no mess and untidiness to find the problem in the field which was In the case of traditional method of farming.

VI. DESIGN PHASES OF PROJECT

Phase 1: Planning, Analysis, Designing and Implementation.

Analysis: Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation.

Planning: After analysis we will first study about it and do some research on it for our better understanding of the project and also get a rough picture about what would be our problem definition for the particular project.

Designing: Then we will construct the design of the project and according to that, will list down all the requirements needed for the construction for the prototype of our project.

Implementation: After acquiring the requirements we first developed the connections of the hardware components (Arduino, LEDs etc) and then added the coding part later with help of Arduino data cable to the Arduino.

Phase 2: Testing and Deployment

Testing: After the prototype is ready we will first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding to it and will check again. Deployment: After, complete integration and testing of project real time running and operation of the system will be done. User need to check the value.*Subject Learning Outcomes and Contents*

VII. ANALYSIS AND PLANNING

A. Feasibility Study

Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation. After analysis we will first study about it and do some research on it for our better understanding of the project and also getting a rough picture about what would be our problem definition for the particular project.

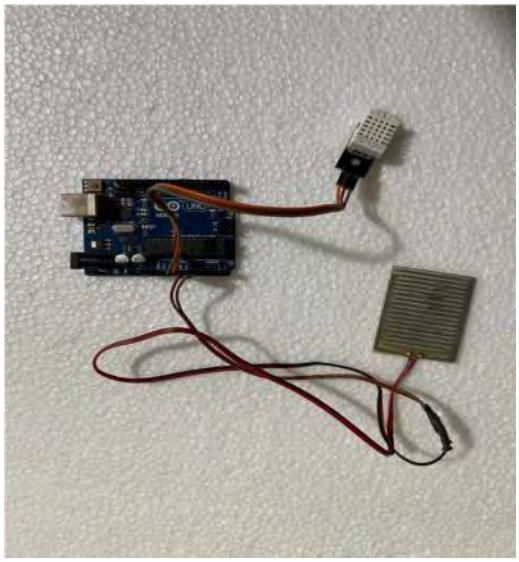
Technical Feasibility: Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

Economic Feasibility: As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

Smart Wardrobe

B. *Cost estimation:*

Components	Cost
Arduino uno, jumpers, resistors, led,bread board	Rs. 720
DHT22 Temperature and Moisture sensor	Rs. 250
Total	Rs. 970



C. Project planning (Resources,Tools used, etc.) Hardware Tools:

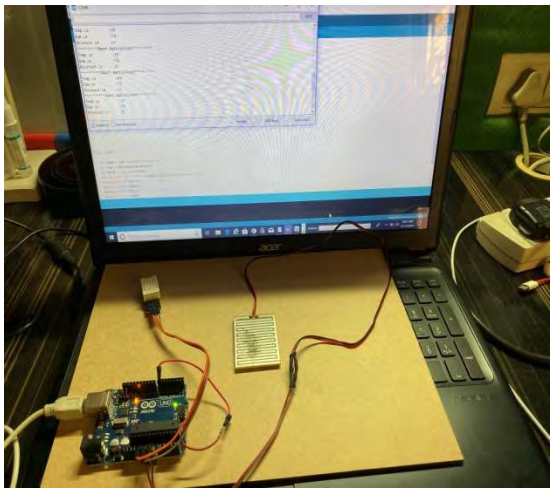
Hardware Components

1. Arduino Uno
2. Breadboard
3. DHT22
4. Connecting wires

D. Software Tools:

1. Arduino uno

E Result(Expected Ourtcome):



VIII. CONCLUSION

IoT-Based temperature and humidity detecting system provide an efficient and definitive system for monitoring agricultural parameters. The corrective action can be taken. IoT-Based monitoring of field not only allows user to reduce the human work and time, but it also permits user to analyze accurate changes in the atmosphere and for taking possible action. It is cheaper in cost and consumes less power. The GDP per capita in agro sector can be increased. This IoT-based system can be extended for controlling different electronic and electrical apparatus from remote locations and the system can also extended for soil moisture and cattle monitor.

IX. ACKNOWLEDGEMENT

We would like to thank subject lecturers (current and previous), education designers and other staff member from the Thakur College of Engineering and Technology who previously designed, developed and taught this subject at various times. The success and final outcome of this project required a lot of guidance and assistance from many people and we am extremely privileged to have completed the project successfully. We would like to thank everyone for their guidance.

References

[1] Adomavicius, G. and Tuzhilin, A. (2001) 'Using data mining methods to build customer profiles', Computer, Vol. 34, No. 2, pp.74–82.

[2] Amato, G. and Straccia, U. (1999) 'User profile modeling and applications to digital libraries', Proceedings 3rd European Conf. Research and Advanced Technology for Digital Libraries, Springer-Verlag, pp.184–197.

[3] Balabanović, M. and Shoham, Y. (1997) 'Fab: contentbased, collaborative recommendation', Communications of the ACM, Vol. 40, No. 3, pp.66–72. [4]Chen, H.C. and Chen, A. (2001) ' Conference on Information and Knowledge Management,from the Thakur College of Engineering and Technology who previously designed, developed and taught this subject at various times. The success and final outcome of this project required a lot of guidance and assistance from many people and we am extremely privileged to have completed the project successfully. We would like to thank everyone for their guidance.

[4] Adomavicius, G. and Tuzhilin, A. (2001) 'Using data mining methods to build customer profiles', Computer, Vol. 34, No. 2, pp.74–82.

[5] Amato, G. and Straccia, U. (1999) 'User profile modeling and applications to digital libraries', Proceedings 3rd European Conf. Research and Advanced Technology for Digital Libraries, Springer-Verlag, pp.184–197.

[6] Balabanović, M. and Shoham, Y. (1997) 'Fab: contentbased, collaborative recommendation', Communications of the ACM, Vol. 40, No. 3, pp.66–72. [4]Chen, H.C. and Chen, A. (2001) ' Conference on Information and Knowledge Management,

Smoke Detector using Arduino

Vivek Mishra, Ritesh Mishra, Ashish Mishra
TCET, MUMBAI
vm105224@gmail.com, imritesh198@gmail.com, vibhormishra98.vm@gmail.com

Abstract— **Over the years, IoT, i.e. Internet of Things, is the currently the most trending field in the Branch of Information Communication Technology. In case of construction areas, residential and commercial buildings, there are high chances of domestic fire. Moreover, mining industry being the most fire prone area, there is requirement of a system that can generate information of the region to the miner on the basis of the gas getting generated while mining. Gas is generated first and then fire. The objective of this project is to build a smoke detection system based on IoT technology. A smoke detector is a device which senses the gas and based on the Gas Quantity that is present in the atmosphere it will provide alerts to the beholder.**

Nowadays, Gas Detection System has been the basic requirement in various regions as the damage caused by the fire accidents is catastrophic. Also, as now the apartments are pretty close and congested, fire can spread very quickly. So, if gas detection and system is Installed while building the apartment, the fire can be suppressed while it is basically in the Small area and further loss of life and property due to fire can be prevented. After studying the literature survey, it was planned to overcome the research gaps and the system was tested using the test cases and then the outputs were generated. The output of the gas detectors were been shown according to different test cases and thus the gas detection system become more responsive then those currently in use.

Keywords – Smoke Detector; LCD screen; Mobile Application; Arduino MQ135

I. INTRODUCTION

Commercial buildings, Mining Regions, Factories, etc. Nowadays has become more prone to fire because of some or the other issues like short circuit, gas leakages etc. So, there is a need of Gas detection system that can help us to overcome the damage to the lives and valuables. It is the fact that Gas is the first to come and then the fire. Hence detecting the fire at the gas level can help avoiding the fire to cause catastrophic destruction. The Gas detection a system is a Device that detects relevant amount of gas in the region and based on that it generates the output that there is a fire or not and based on the output of the further actions by fire-fighters and the administrations can be taken. Moreover now a days there has been an increase in such cases these days, this project also take an initiative to overcome that by discussing about it in the future

Gas leakage detection has become a crucial aspect in design of buildings, both commercial and domestic, as opposed to about 70 years ago when automatic detection was rarely provided in buildings. Before introduction of gas alert, fires resulted in the loss of human lives and damage of property and it was mainly attributed to lack of a mechanism for early detection of gas. Early developments in design of gas alarm began in 1922 with observations by Greinacher and later by Walter Jaeger in 1930. However, early gas detectors required high voltage power input. Further research was done and power requirement in gas detectors was reduced to make battery power viable and this made widespread installation in residences highly feasible. Later developments in gas detectors have sought to improve their performance, reduce power requirement and improve their nuisance alarm sensitivity and also to continuously monitor their status. The most recent advances in gas detectors have been motivated to make them “smarter”. Gas can be detected either optically (photoelectric) or by physical process (ionization). Detectors may use either or methods. Subject

II. SUBJECT HISTORICAL DATA

In this section, we present the previous research done by the people by reading their papers published. Also, the problem definition, features of the project and methodology is explained and learned in detail.

A. Literature Survey

A literature review surveys scholarly articles, books, dissertations, conference proceedings and other resources which are relevant to a particular issue,
area of research, or theory and provides context for a dissertation by identifying past research.

Research tells a story and the existing literature helps us identify where we are in the story currently. It is
up to those writing a dissertation to continue that story with new research and new perspectives but they must first be familiar with the story before they
can move forward.

B. Problem Definition

We need to design a fire alarm system that all family members can use in single-family residences. It must be able to detect fires at all locations, residents must be able to activate it from convenient locations themselves, and it must alert residents in all portions of the house. This system should reduce the cost of fire-insurance protection as well as provide security for the homeowner. A new housing development opening in 12 months is going to include this product in a list of
options for homeowners to purchase when finishing their new home.

C. Features of the Project

Smoke Detectors are very useful in detecting smoke or fire in buildings, and so are the important safety parameters. In this DIY session, we are going to build a Smoke Detector Circuit which not only sense the smoke in

the air but also reads and displays the level of Smoke in the Air in PPM (parts per million). This circuit triggers the Buzzer when Smoke level becomes higher than 1000 ppm, this threshold value can be changed in the Code according to the requirement. This circuit mainly uses MQ135 Smoke/Gas sensor and Arduino to detect and calculate the level of smoke. MQ 135 gas sensor is also sensible to LPG, Alcohol, and Methane etc.

The features of our project can be highlighted in following points:

•Simplified automation.
•Minimal controls
•Cost efficient
•Can be upgraded if needed

D. Methodology Used

The proposed method takes an automatic control action upon detection of gas. Initially if there is a Leakage detected then the electronic sensor i.e. the gas sensor that obeys the principle of sensor senses any fire in the region, if any leakage or fire is sensed then the output of this sensor goes high. This high signal is monitored by the microcontroller and it will identify the hike in the leakage amount. If there is a leakage or smoke detected, the consumer is informed about it through the LED light and Sound of the buzzer

II. ANALYSIS AND PLANNING

Analysis is an important part of any project to know calculate the wrongs and rights of the project i.e. the economic study tells us the capital used, the planning tells us about the time taken, resources used gives us all idea about how well the project is being going on.

A. Fesiability Study

Analysis is an important part of any project to know calculate the wrongs and rights of the project i.e. the economic study tells us the capital used, the planning tells us about the time taken, resources used gives us all idea about how well the project is being going on.

1. Economic Feasibility:

As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So, the proposed system is economically feasible.

2. Technical Feasibility:

Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

3. Operational Feasibility:

It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.

B. Assumption Risk:

a) Assumption:

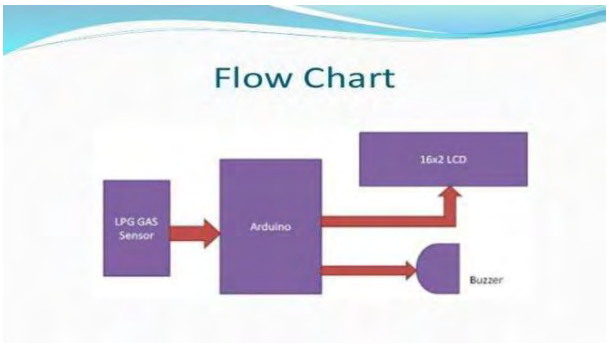
Management will ensure that project team members are available as needed to complete project task and objectives. The team members will participate in the timely execution of the Project Plan (i.e., timely approval cycles and meetings when required.)

b) Risks:

Accurate identification of the system requirements, interpretation and design depend on the patience and commitment of the team members in focus.

IV. DESIGN REQUIREMENTS

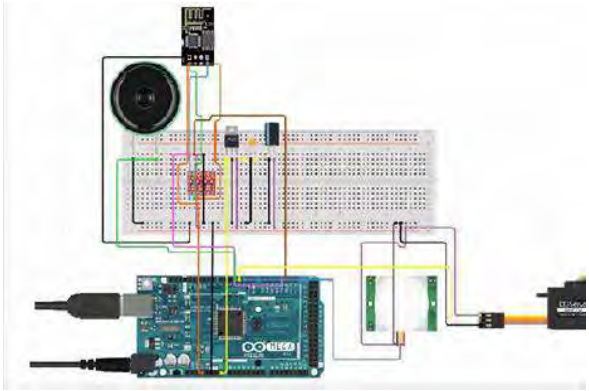
C. Process Flow



This is the process flow of the project.

D. Block Diagram

The following is the Breadboard implementation of the project which can also be converted to a PCB in order to make the assembly smaller.



V. RESULT

OUTPUTS:

When the smoke level was below the threshold value the environment is safe and LCD shows the respective status and when high makes a beep sound.

VI. CONCLUSION

We have successfully shown the detection of Gas Leakage by IOT based project is possible which can detect Gas leakage and warn owners and remote contacts of a alarm situations. The alarms are detecting reliably even in situations where there is high ambient noise. A weakness of the system is in the processing overhead used to constantly monitor the audio input of the phone. This application demonstrates the mobile phones ability to be used as a advanced warning device in the event of a smoke alarm activation, its automated group alert system reduces the chance of harm to people who have not been able to respond to an alarm.

VII. FUTURE SCOPE

Sometimes it happens that fake smoke alarms are generated which can cause a loss of more than $12,000 US dollars. So to avoid and tackle that we can use the camera and delay by 30 second in alarm which means if the smoke is detected to be above the threshold quantity, the detector will send the video to the operator and if the operator approves that it is a legitimate fire in the region, the alarm will be horned and within 30 seconds of no reply from the user, the alarm will ring and accordingly the further actions can be taken in future.Gas Leakage detection and alarm system can be used with quad copter and this quad copter can be used to detect the regions of the fire or going to catch the fire, thus can be used to protect things.

VIII. ACKNOWLEDGEMENT

We sincerely thank our guide Mr. Anil Vasoya for his guidance and support for carrying out our project work. We also thank all the project coordinators for arranging the necessary facilities to carry out the project work.

We thank our Principal, Dr. B.K Mishra, HOD, Mr. Rajesh Bansode and the College Management for their support. We would like to thank all our classmates & faculty members for their inputs and valuable insights.

References

Journal Refrences:

[1 douglas mr, mallonee s, istre gr. estimating the proportion of homes with functioning smoke alarms: a comparison of telephone survey and household survey results. am j public health 1999;89:1112-1114. [pmc free article] [pubmed] [google scholar]

[2] ahrens m. us experience with smoke alarms and other firealarms: who has them? how well do they work? when don't theywork? quincy, ma: national fire protection association;1998.

[3] smith cl. the smoke detector operability survey: report on findings. bethesda, md: consumer product safety commission;1994.

[4] shults ra, sacks jj, briske la, et al. evaluation of three smoke detector promotion programs. am j prev med 1998;15:165-171. [pubmed] [google scholar]

web references:

1.http://www.swsciences.com/research/papers/ch enfiredetection2007.pdf

2.c:\sjchen\publications\submitted\firesafetyjournal 2004\finalized_manuscript_u161_january2007.w pd

3.http://www.swsciences.com/research/papers/ch enfiredetection2007.pdf

4.c:\sjchen\publications\submitted\firesafetyjo urnal-2004\finalized_manuscript_u161_january2007.w pd

5. https://www.oaktreeproducts.com/smoke- detector-research

IoT based Child Tracking Security

Rachita Mishra, Anushka Puranik, Kimaya Patil
Department of IT Thakur College of Engineering and Technology
rachitamishra.rsanp@gmail.com, puranikanushka@gmail.com, kimayapatil427@gmail.com

Abstract—**The era of the Internet of Things (IoT) has already started and it will profoundly change our way of life. While IoT provides us many valuable benefits, IoT also exposes us to many different types of security threats in our daily life. Before the advent of IoT, most security threats were just related to information leakage and the loss of service. With IoT, security threats have become closely related to our non-virtual lives and they can directly influence physical security risk. The Internet of Things consists of various platforms and devices with different capabilities, and each system will need security solutions depending on its characteristics. There is a demand for security solutions that are able to support multi-profile platforms and provide equivalent security levels for various device interactions. In addition, user privacy will become more important in the IoT environment because a lot of personal information will be delivered and shared among connected things. Therefore, we need mechanisms to protect personal data and monitor their flow from things to the cloud.**

I. INTRODUCTION

Child safety and tracking is a major concern as the more number of crimes on children are reported nowadays. The chance of missing children is going in rampant. With this motivation, in order to reduce this problem, a smart IoT device for child safety and tracking is developed using Arduino Uno and GPS tracker so that parents can be able to locate and monitor their children.

This system is used for tracking the information of the lost child using Google Map along with the position and location of that child through GPS.

This process operate simply by keeping the "tracking system device" into the bag of that particular child, who is going to school or outside world and now if in case that child is lost or missed then the parents of that particular child can simply track him/her by the device which has been kept inside that child bag.

In this way the parents get the real time location by receiving the exact position of the child along with the longitude and latitude of that place then it will be copied into the Google map and the location of that lost child can easily be accessed. The above system ensures the safety and

Tracking of children.

Now a day's child tracking system is widely used all over the world and it gives the assurance to the parents that their child is safe from suspicious action. In this project we have shown the system requirement for tracking the child and describe the implementation feature.

To implement such system a GPS with high accuracy is required, if the low accuracy GPS is used in this system. System may give some error of that child location. In this system GPS of that device send the longitude and latitude to the GSM module, GSM module receive the information about longitude and latitude of that child location, further this message will be sent to the parents for tracking the location of that lost child. The Arduino is a small micro controller which is used for controlling whole process in this system. This project is for developing a low cost, high accuracy and user-friendly system by using Google map. Google map can improve the accuracy of GPS. Improvements are proved by Google map that make high accuracy.

II. BACKGROUND

People cannot predict or determine what will happen next to them in next seconds. This is the same scenario as people cannot acknowledge of when will be targeted, but can make precaution prevent it from happening. As for the great parents, their greatest concerns are their child, especially regarding the safety of the child.

The parents start to take care of the child since born. The parent is not possible to always stay beside of them as most of the parents needs to go for work to earn some money to sustain a family. Parents will start to feel anxious about their child current status every second because they cannot see what their child is doing currently or what affairs will happen on them.

Today's child is easier influenced by their friends, and they might even get cheated or kidnapped by any of the strangers, as what had happened to a girl, named Nurlin Jazlyn, who had gone missing on August 20, 2007 after she went alone at a night market in Kuala Lumpur. She was found dead in a sports bag after a month of investigation by police (Theage.com.au, 2016).

To minimize this tragedy from happening again, an action needs to be taken to deal with the problem. Thus, in order to resolve these problems the child tracking system is proposed..

III. SCOPE OF PROJECT

This project is actually developed for parents to keep track the child whereabouts. Nowadays, child is easier influenced by their friends, and they might even get cheated or kidnapped by any of the strangers. By developing this system can track child current location.

The application is mindful to keep track the location of the device. In this application the parent track for the route their child traversed during a certain period of time. The application in the device will give the real time location of the child. Parents can take immediate action if any inconvenience happens when the location not found or track.

IV. SUMMARY

The effectiveness of this project depends on how the products are detailed and how the components are used. The most important part is of the GSM module and GPS tracker as GSM is used for communication and GPS is used to get the location and if it works to fail then it is not possible that the model can work. The cost is also low as per our model requirements.

V. LITERATURE SURVEY

So during literature survey we came across few projects where the first one named "Child Tracking System" was about 2 users' namely parent and child. The parent sends a request to the server for the location of the child, and the server will send the request to the child app about the request. When the request received and location retrieved, the child app will send the request again to the server about the newly detected location. The server then will send the details to alert the parent about the new location and show the latest location inside the map of the app.

The second was similar namely "Smart IoT device for Child Safety and Tracking" which proposes Child safety as GSM module is used. Parent will send message to GSM module. According to message received GSM module to reply back with particular details of the child. The location can be seen on Google map. When a child is in any emergency situation, device button should be pressed so that the device captures the image along with the user information to the enrolled mobile number. The life of children can be saved within no time.

The following paper "IoT based Unified approach for Women and Children Security using Wireless and GPS" presents a wireless method which will alert and communicate with secure medium. It will also capture image via camera. When the sensor kit button is pressed the camera will captures the image and will collect the information of the user. This information will be sent to the registered phone number along with the image link. Children security can also be done by using the GPS tracking mechanism. This Unit will locate the bus and all its travelling routes. Alert messaging will be done on the registered phone numbers.

The last was "IoT based Smart GPS Device for Child and Women Safety Applications" in which an Android based solution to aid parents to track their children in real time. The concerned device is connected to server via internet. The device can be used by parents to track their children in real time or for women safety. The proposed solution takes the advantage of the location services provided by GSM. It allows the parents to get their child's location on real time by SMS.

VI. PROBLEM DEFINITION

1) Increasing rate of child kidnapping.
According to News Strait Times Newspaper, of the 567 children, another 193 children are still missing and being investigated. A majority of cases were due to family problems, seeking freedom and running off with lovers or friends. (Bukit Aman CID (Investigations and Legal) deputy director Datuk Law Hong Soon, June 4, 2016)

2) Lack of tracking technology for child.
The parent is hardly to keep a watch on their child without the use of technology, especially when the child is in the outdoor. The parent even cannot avoid the negligence that will make by children in the future day.

3) Limited application for child monitoring.
There is very limited application available for tracking child when they are out of parents control and let kidnapping or missing cases occurred.

VII. FEATURES OF PROJECT

The proposed system can reduce the child missing cases. This system provides tracking solution for the parent to keep tracking their child location in the outdoor by using GPS as where it allows determining the exact location of the child. Moreover, help to minimize this tragedy from happening again and again in future. This project is simple and cost efficient.

VIII. METHOLOGY USED

As crime's are increasing day-by-day safety for all children have become a concern for all parents. In this project we have use GSM module and GPS tracker which will help to track the exact location of child in real time.

1X. DESIGN RQUIREMENT

A) BLOCK DIAGRAM

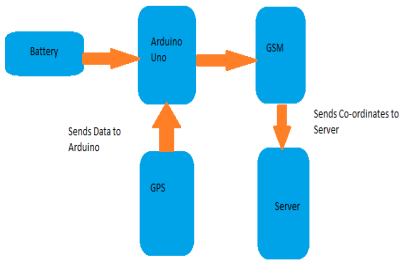


FIG 1X.1 block diagram of IoT based Child Tracking System.

B) SCHEMATIC DIAGRAM

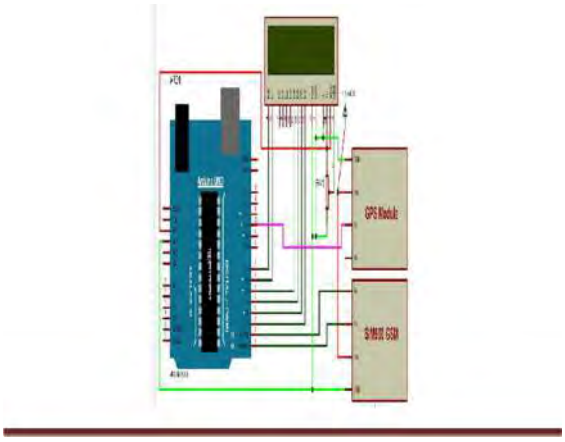


FIG IX.2 schematic diagram/ circuit diagram showing all the necessary connection that need to be made.

X. RESULT

Results show that everything worked as anticipated, although some high pings in the network may cause minor occasional lag in the connection and data exchange, however, there was no major lag detected affecting the model control and testing. The lag maybe a casual effect of using 3G connection. This project was successful to track and get the current location on mobile phone.

- When Location is tracked SMS is send on parents' mobile phone which contains coordinates of current location of child
- After clicking on the SMS link a google map is open in which location can be seen.

FIG X.1 Tracked Location

FIG X.2 Parent Message Screenshot

XI. FUTURE SCOPE

This research demonstrates Smart IoT device for child safety and tracking helping the parents to locate and monitor their children. If any abnormal values are read by the sensor then an SMS is sent to the parents mobile and an MMS indicating an image captured by the serial camera is also sent. The future scope of the work is to implement the IoT device which ensures the complete solution for child safety problems.

XII. CONCLUSION

The child tracking system has been experimentally proven to work satisfactorily by trying to track the current location using GPS and see the location on mobile phone.
Hence, the idea to create a fully working child tracking system with facility to track the location has been finally realized. The output of this project is that, this project can be of huge help to parents where security of children is concerned.
This project is a successful outcome of continuous and tireless effort from all the project members, supervisors, college faculty, colleagues and other helping hands. Thus, a low-cost child tracking system was successfully designed, implemented and tested.

References

Journal References:

[1] Fairuz Rauf, Zuraidy Adnan, Go this wary Subramaniam. "Child Tracking System". International Journal of Computer appliances, Volume 181, No.3, July 2018.

[2] M Nandini Priyanka, S Murugana, K N H Srinivas, T D S Sarveswarao, E Kusuma Kumari. "Smart IoT device for Child Safety and Tracking". International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8 Issue-8, June2, 2019.

[3] Deepali Bhavale, Priyanka Bhavale, Tejal Sasane, Atul Bhawane. "IoT based unified approach for Women and Children Security using Wireless and GPS". International Journal of Advanced

Research in Computer Engineering and Technology(IJARCE), Volume-4, Issue-3, May-June 2016

[4] Gaikwad Priyanka Rajaram, Gotraj Sonali Machindra, Jagtap Pooja Implementation Remdas , Pagare Prajakta Yaswant , Review on of Child Tracking System Using Mobile ,Imperial Journal of Interdisciplinary Research , Vol2, Issue4, 2016

[5] Al-Mazloum, A., E. Omer, and M. F. A. Abdullah. "GPS and SMS-based child tracking system using smart phone." International Journal of Electrical, Robotics, Electronics and Communications Engineering 7.2 (2013): 171-174.

Mobile Wireless Technology Evolution from 0G to 5G

Pratimesh Majumdar, Pradum Chauhan, Pratik Kumbhar
TCET, MUMBAI
pmajumdar912@gmail.com, chauhanpradum1681@gmail.com, pratikkumbhar121999@gmail.com

Abstract- **In wireless communications, the world is undergoing a major technological resolution that will provide ubiquitous access to communication to people wherever they are. The exponential evolution of wireless technologies, services and enterprise applications has resulted within the deployment and use of wireless and mobile networks on an outsized scale. The mobile wireless advanced in a totally quick span of time. Wireless access technologies are about to be built to hit its fifth generation (5G). Looking back, wireless access technologies have taken different evolutionary paths towards a single goal: performance and productivity in highly mobile environments. The first generation fulfilled the simple mobile voice while capacity and coverage were added by the second generation. This was followed by the third generation which aims at higher speeds fir data, The fourth generation provides access to a wide variety of telecommunication services which are supported via mobile and fixed network. This paper provides an outline of the evolution of Mobile Wireless Communication Networks from 0G to 5G.**

Keywords – wireless, Mobile wireless communication, 0G, 1G, 2G, 3G, 4G, 5G,GSM

I. INTRODUCTION

Nowadays, the wireless technology is getting popular and important within the network and in Internet field. Mobile communications are clearly showing major enhancements in terms of capabilities of mobile networks. the following generation of wireless services, besides improving the general capacity, will create its own unique demands in terms of localization, personalization, etc., which is able to successively, drive the event and continuous evolution of services and infrastructure. In this paper we briefly introduced the history background of 0G to 5G, compared them, and illustrated how 5G may match for more convenient and powerful in the future. There are many improvements from 0G,1G, 2G, 3G, and 4G to 5G in the world of telecommunications. Evolution of 4G is based on the limitations of 3G and it fulfils the amount of an inspiration that's WWWW, World Wide Wireless Web, offering more services and smooth global roaming with inexpensive cost.

II. EVOLUTION OF MODILE CELLULAR NETWORK

Zero Generation Telecommunication (0G) :

0G – Mobile radio telephone (also known as "0G"). 0G alludes to pre-cellphone portable communication. These frameworks are called 0G (zero age) frameworks. Advances utilized in 0G frameworks included PTT (Push to Talk), MTS (Mobile Telephone System), IMTS (Improved Mobile Telephone Service), and AMTS (Advanced Mobile Telephone System). Since they were the antecedents of the original of cell phones, these frameworks are once in a while retroactively alluded to as pre-cell (or some of the time zero age) frameworks.

1G Technology :

These phones were the first and the premier cell phones to be used. In 1979, 1G was propelled economically in Tokyo city of Japan by NTT DoCoMo firm.

1G is a simple innovation and the telephones commonly had poor battery life and voice quality was huge absent a lot of security, and would frequently encounter dropped calls. The most extreme speed of 1G was 2.4 Kbps.

1G innovation wasn't completely analog. Information transmissions were done in simple structure at 150 Mhz or more recurrence of radio waves. This was the greatest downside of 1G innovation. This prompted instability of calls. Likewise, because of less recurrence transfer speed for 1G, information transmission speed was excessively low and was attainable for calls as they take low information to convey the expected signs.

Call was the main medium to pass on your message to the next individual. Information transmission speed was around 3-4 kbps. Because of low-recurrence transfer speed and simple type of information transmission, it used to take bunches of vitality to transmit information to far separations and furthermore signals would get upset while arriving at their specific beneficiary. Information transmission used to take loads of vitality. The other large downside was roaming that was not upheld in 1G innovation.

Downsides of 1G:

- Poor voice quality
- Poor battery life
- Large telephone size
- No security
- Limited limit
- Poor dependability

2G Technology :

2G or second era of remote innovation was first propelled financially in Finland in 1991. It used analog signals for voice transmission. Crucial focus of this innovation was on advanced signals and offered administrations to pass on content and give picture messages at low speed (in kbps). It used the data transmission of 30 to 200 KHz. This gave us smaller gadgets, a progressively secure association, better call quality, and a higher limit with regards to network.

During the 2G period, there were many progressions made inside the range itself, for example, GSM, GPRS and EDGE.

- GSM: Short for Global Systems for Mobile Communication empowered information move over voice correspondence at speeds around 30-35 kbps.
- GPRS: General Packet Radio Service worked on the comparable 2G innovation as GSM with some refinements which gave it higher information speeds (110 kbps) .

• EDGE: Enhanced Data rates for GSM Evolution presented in 2003 was to some degree known to be 2.9G or 3G because of its noteworthy headways over GPRS and GSM. It offered high speeds of 135 kbps.

2G: SMS and MMS

• The 2G phone innovation presented call and content encryption, in addition to information administrations like SMS, picture messages, and MMS.
• The max speed of 2G with General Packet Radio Service (GPRS) is 50 Kbps or 1 Mbps with Enhanced Data Rates for GSM Evolution (EDGE).
In 2G, three major highlights were added and improved contrasted with 1G and they are:

1) Data transmission for telephone discussion got completely digitized and encoded.
2) Internet information administration and SMS like plain content, picture messages, MMS administration were started notwithstanding calls.
3) 2G empowered gadgets turned out to be progressively effective to get and impart signs with higher infiltration levels for wide radio waves range.
Better voice quality, more than one medium to convey acquired another change remote correspondence innovation. With information encryption strategy, transmission turned out to be increasingly productive and quick with better quality and information security. Just a specific beneficiary could get the proposed message by means of encryption procedure.
The digitized transmission prompted advancement of littler mobiles since simple parts were disposed of with this age. The battery life improved because of digitization of information transmission which normally took less vitality to transmit. This occurred because of two diverse multiplexing principles on which 2G innovation depended on : GSM (Global Systems for Mobile) in light of TDMA (Time Division Multiplexing Access) and CDMA (Code Division Multiplexing Access). This additional the circuit exchanging area which permitted a solitary association progressively. On the off chance that the system bounced from one base station to other base stations, at that point the association would in general disengage and everything must be re-started. Information transmission speed was additionally expanded to 14.4 kbps contrasted with 3-4 kbps in 1G.
Downsides of 2G:
• Required solid advanced signs to enable cell phones to work. On the off chance that there was no system inclusion, advanced signs debilitated.
• Systems couldn't deal with complex information like recordings.

2.5G Technology :

2.5G was an intermediate stage between 2G and 3G innovation. Packet switching domain was included. This empowered the sending of any solicitation to arrange bases as little packets as opposed to a real-time connection. It assisted with sparing battery much more as the transmissions were in little packet structure. This forestalled the need to retransmit the information.

2.5G or GPRS image :

2.5G speaks to handsets with information capacities over GPRS. Packet exchanging space sped up to 50-60 kbps than in typical 2G innovation. This started web convenience and furthermore presented GPRS(General Packet Radio Service). GPRS was indicated utilizing G symbol.

2.75G Technology :

2.75G or EDGE (Enhanced Data rates for GSM Evolution). 2.75G was improved GRPS frameworks to EDGE (Enhanced Data Rates for GSM Evolution) frameworks with 8PSK encoding strategy. This improved the information rates by 3-4 times. In 8PSK encoding, a solitary transporter image used to convey 3 bits rather than 1 piece utilized in GPRS. This further improved voice call quality. Information transmission hypothetically expanded to 1Mbps (down to earth assessed speed is 500 kbps). EDGE or 2.75G is generally indicated by the E symbol close to the versatile system.
EDGE innovation is an all-inclusive form of GSM. It permits the unmistakable and quick transmission of information and data. It is likewise named as IMT-SC or single transporter. EDGE innovation was concocted and presented by Cingular, which is currently known as AT& T. EDGE is radio innovation and is a piece of third era advances. EDGE innovation is favored over GSM because of its adaptability to convey packet switch information and circuit switch information.
EDGE moves information in less seconds in the event that we contrast it and GPRS Technology.

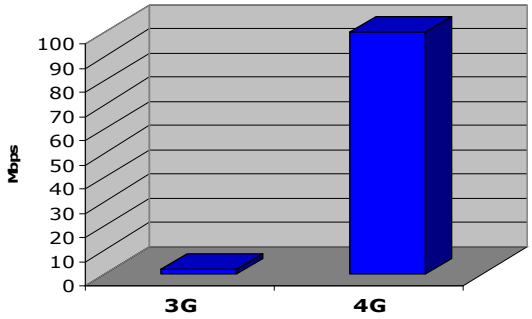
3G Technology :

3G systems were presented in 1998 and represent the cutting edge right now. 3G introduced quicker information transmission speeds empowering video calling and internet usage.
The maximum speed of 3G was assessed to associate with 2 Mbps for immobile gadgets and 384 Kbps in moving vehicles. The hypothetical max speed for HSPA+ is 21.6 Mbps.
3G had and has speed abilities of up to 2 mbps. It empowered cell phones to give quicker correspondence, send/get huge messages and messages, give quick web perusing, video gushing and greater security among others. It was generally founded on CDMA2000 (Code-division numerous entrance) and EDGE advances. It empowered numerous channels to impart at one same along these lines ad libbing on the over speed and availability.
The primary qualification somewhere in the range of 2G and 3G that permitted media spilling to happen is that 3G uses packet exchanging information transmission as opposed to circuit exchanging. Information is separated into little pieces or parcels and afterward sent to the goal. Utilizing this technique for transmission extraordinarily speeds up, permitting one to send information through various diverts in equal instead of one direct in arrangement. This innovation additionally permits clients to pay for information utilized as opposed to time spent on the web.
3G innovation gives better voice quality and better availability because of wide recurrence transmission capacity. W-CDMA standard is broadly utilized in 3G

innovation as it gives preferred information transmission speed over UMTS.
3G innovation upgraded information transmission and offered better information rate in any event of 384kbps for moving position and 2Mbps for stationary position. Quicker information rates and availability permitted organizations to grow. Better scrambling/encoding strategy used to transmit information included a superior information security. 3G gadgets decreased in size and furthermore battery life improved with better recurrence and encoding strategy.

3.5G or HSPA image :
3.5G or state HSPA (High Speed Packet Access) was an upgraded form of WCDMA innovation dependent on GSM models. Hypothetical information rate was improved to 14Mbps at the uplink/transfer side and 5.76Mbps at the downlink/download side.

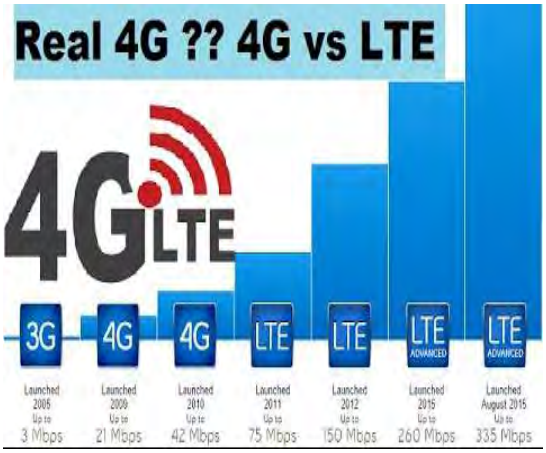
4G Technology :
4G network is the further evolution in the wireless network aiming to provide much faster ans secure network than the generations before as it aimed towards better reception, faster information and data exchanges and less dropped data.
4G simply means "fourth generation". The first generation 1G came in existence 1981 where the means of transmission was analog, and 1992 saw the transformation to 2G where the means of transmission was digital. 3G first came into use in 2001 and included different multi-media support along with highest data transfer rate of at least 200 kilobits per second. It thus justifies, that 4G means "fourth generation" and represents a number of advancements and enhancements over the 3G technology.



4G POSSIBILITIES :
- Experience improved remote capacities that convey portable gaming cooperation with latency under five milliseconds.
- Play online multiplayer games while going at high speeds or sitting outside.
- Individual Media Repository.
- Make an individual media vault that can be gotten to from home and on the road to see photographs, watch films and tune in to your own music choices
- Broadband Access in Remote Locations
- 4G systems will give a remote option in contrast to broadband access to private and business clients.
- Also, 4G will give the first chance to broadband access in remote locations without a framework to help link or DSL get to.

Table 1 3G vs 4G

Technology	**3G**	**4G**
Data Transfer Rate	3.1MB/sec	100MB/sec
Internet services	Broadband	Ultra Broadband
Mobile-TV Resolution	Low	High
Bandwidth	5-20 MHz	100+ MHz
Frequency	1.6-2 GHz	2-8 GHz
Network Architecture	Wide Area Network	Hybrid Network
Signal Quality	Good	Best
Communication Type	Circuit Switching	Packet Switching



4G Access Options:
- Making end user devices 4G compatible.
- Develop mobile personal gateways .
- Develop external interfaces with 4G peripherals.

4G advantages:
- High voice quality.
- High spectral efficiency
- Easily access internet, streaming media, video calling etc.
- Very low latency.
- Simple protocol architecture.
- Efficient multicast/broadcast.

Disadvantages of 4G:
- New frequencies means new components in cell towers.
- Higher data prices for consumers.
- Consumer is forced to buy a new device to support the 4G.
- It is impossible to make your current equipment compatible with the 4G network.

5G (5th generation) :

Problematic advances, for example, independent vehicles and Internet of Things (IoT) in the present occasions have given a driving force for 5G innovation. For instance, DHL's IoT following and checking framework tracks everything from vehicle conduct to bundles to natural sensors in the airways . In this way, information volume and peak rates are going increment over the wireless transmissions. Thus, the portable correspondence industry is thinking of 5G arranges as a successor to 4G.

Over the most recent three years, the 5G systems have begun getting footing with institutionalization endeavors at European Telecommunications Standards Institute . Sooner, portable administrators in the European Union (EU) and USA began certifiable preliminaries of 5G systems. With 5G, accelerate to 10 Gbps can be accomplished.

Accordingly, a portion of the application regions where 5G assumes a noticeable job incorporate Smart Homes, self-driving autos and VR/AR applications.

5G remote system or 5G net. As we probably am aware, 2G (second Generation) remote system began in the mid 1980s and fundamental telephone with voice and information administrations was advertised. Next, 3G (third era) remote system were presented offering information move with max download speeds going up to 4.4 Mbps for HSPA and 168 Mbps for HSPA+. With 3G, video calling administrations were empowered.

Here are some challenging elements which should be tended to before we get 5G out into the normal man's hand:

• At this point, most geo-areas are not open enough to oblige 5G receiving wires. The earlier predominant radio wires' set up should be moved up to convey new ones that will keep up the 5G range.

• There is a worry with respect to the Clouding of the Air Interface and range movement. Be that as it may if Air Interface is 5G Clouded, powerful allotment of 2G, 3G, 4G, and 5G web will be allowed, and that thus will guarantee help of a smooth procedure to the extent relocating the range.

• Since undoubtedly, the Distributed Antenna System (DAS) – the one we have right now – doesn't flaunt enough ability to help 5G, to satisfy this interest a flat out reestablishment of the current advanced framework will be required if without hiccup migration of 5G is to be directed throughout the following scarcely any months.

	4G-LTE	**5G**
Download real-world	42 Mb/s	100 Mb/s
Uplink real-world	25 Mb/s	50 Mb/s
Download Theoretical	1 GB/s	20 Mb/s
Uplink Theoretical	500 Mb/s	10 Mb/s
Spectral Efficiency (Downlink)	15 bps/Hz	30 bps/Hz
Spectral Efficiency (uplink)	6.75 bps/Hz	15 bps/Hz
Latency (Control plane)	100 ms	50 ms
Latency (User plane)	10 ms	1 ms
Frequency Bands	700-2100 MHz	28-40 GHz

III. CONCLUSION

Wireless technology is becoming widespread and important nowadays in both the network and the internet field. In future, new work can be done on these technologies by introducing new features to them and combining two or more technologies which can be used to get better results. In this paper we briefly introduced the history background of 0G to 5G, compared them, and illustrated how 5G may match for more convenient and powerful in the future. There are plenty of improvements from 0G,1G, 2G, 3G, and 4G to 5G within the world of telecommunications. In the novel impending 5G technology is offered in the market in affordable rates, and high peak future and far reliability than its preceding technologies.

References

[1] Vasari Bande, Mounika Marepalli,Leepika Gudur "Evolution of 4GResearch Directions Towards Fourth Generation Wireless Communication", " International Journal of Computer Science and Information Technologies", Vol. 2 (3) , 2011, 1087-1095.

[2] B. G. Evans and K. Baughan, "Visions of 4G," Electronics and Communication Engineering Journal, Dec. 2002.

[3] H. Huomo, Nokia, "Fourth Generation Mobile," presented at ACTS Mobile Summit99, Sorrento, Italy, June 1999.

[4] http://www.4gwirelessjobs.com/pdf/5g-Wireless architecture.pdf .

[5] http://www.5-g.co.uk/

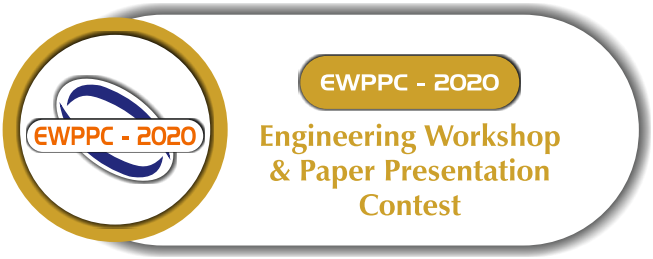
[6] http://5ginfo.blogspot.com/

[7] Rappaport, Theodore S. Wireless communications: principles and practice. Vol. 2. New Jersey: Prentice Hall PTR, 1996.

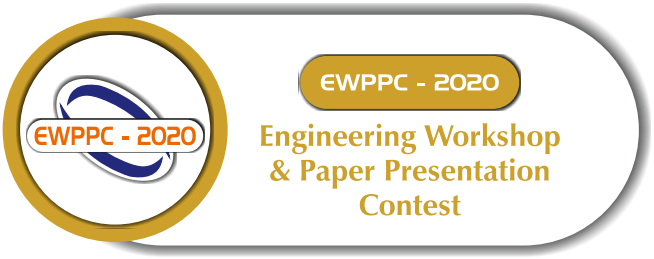
[8] Carney, W. and Solomon, Y. The future of wireless LANs will be multimode. Texas Instruments White Paper SPLA001 (Nov. 2002); focus.ti.com/pdfs/vf/bband/80211_wp_multimode.pdf.

[9] Kamarularifin Abd Jalil, Mohd Hanafi Abd. Latif, Mohamad Noorman Masrek, "Looking Into The 4G Features", MASAUM Journal of Basic and Applied Sciences Vol.1,No. 2 September 2009 .

[10] 3gamericas (2010). "Transition to 4G: 3GPP Broadband Evolution to IMT-Advanced", Rysavy Research/3G Americas. [Online] Available: ww.rysavy.com/PR/3GA_PR_2010_09.pdf.



A - Block, Thakur Educational Campus, Shyamnarayan Thakur Marg, Thakur Village, Kandivali (East), Mumbai - 400 101
Tel.: 022-6730 8000 / 8106 / 8107 Telefax: 022-2846 1890
Email: tcet@thakureducation.org Website: www.tcetmumbai.in www.thakureducation.org



A - Block, Thakur Educational Campus, Shyamnarayan Thakur Marg, Thakur Village, Kandivali (East), Mumbai - 400 101
Tel.: 022-6730 8000 / 8106 / 8107 Telefax: 022-2846 1890
Email: tcet@thakureducation.org Website: www.tcetmumbai.in www.thakureducation.org