

Healthcare Prediction and Analysis System with Constant Data Polling



Thejas Bharadwaj, Aatish Balaji

Abstract: *The applications of Big data and machine learning in the fields of healthcare, bioinformatics and information sciences are the most important things that a researcher takes into consideration when doing predictive analysis. The Data production at this stage has never been higher and it is increasing at an alarming rate. Hence, it is difficult to store, process and visualise this huge data using customary technologies. However, abstract design for a specific massive information application has been restricted. With advancement of big data in the field of biomedical and healthcare domain, accurate analysis of medical data can be proved beneficial for early disease detection, patient care and community services. Machine learning is being used in a wide scope of application domains to discover patterns in huge datasets. Moreover, the results from machine learning drive critical decisions in applications relating healthcare and biomedicine. The transformation of data to actionable insights from complex data remains a key challenge. In this paper we have introduced a new method of polling of data before analysis is conducted on it. This method will be valuable for dealing with the issue of incomplete data and will progressively prompt suitable and more precise data extraction.*

Keywords: *Machine Learning; Data Polling; Data Mining; Healthcare; Customer Relation Management (CRM); Prediction System; Cloud Computing; Data Analysis;*

1. INTRODUCTION

Big Data is defined as a collection of large volume of data that becomes too complex to process by using traditional data processing techniques and platforms. In other words, a data set can be named Big Data if it is difficult to store, process and visualise using basic technologies. Nowadays, data generation sources are increased dramatically, such as streaming machines, high throughput instruments, sensor networks, telescopes, and streaming machines, and these environments produce large amount of data. In recent years, Big Data has been playing a vital role in many environments such as public administration, scientific research, business organisation, healthcare, industry, social networking and natural resource management. Our traditional data processing system and architecture needs to be developed to store such huge size of data. Data Mining is a significant extraction of implicit, previously unknown and potential helpful information about data. In short, it is a process of analysing data from different perspective and gathering the knowledge from it. The discovered knowledge can be utilized for various applications for instance healthcare industry. Nowadays healthcare industry generates a lot of data about disease diagnosis, patients, etc.

Data mining gives a lot of techniques to find concealed information from data. A noteworthy challenge confronting Healthcare industry is quality of service of diagnosis. Quality of service infers diagnosing disease accurately and provide effective treatments to patients. The poor diagnosis can lead to unfortunate results which are not acceptable. Data mining is a fundamental step in the process of knowledge discovery in databases in which intelligent techniques are applied in order to extract patterns. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. The discovered knowledge can be utilized by the healthcare organizations to improve the quality of service. The discovered knowledge can likewise be utilized by the medical professionals to diminish the number of adverse outcomes, to recommend more affordable therapeutically equivalent alternatives. Envisioning patient's future conduct on the given history is one of the vital utilizations of data mining techniques that can be utilized in health care management. For example, it may now be feasible for the doctors to contrast symptomatic data of different patients and indistinguishable conditions. Similarly, doctors can also affirm their discoveries with the help of different doctors working on the same type of domain with an identical case from everywhere throughout the world. Medical diagnosis is considered as a noteworthy yet unpredictable work that should be done accurately and productively. The automation of the same can be proved very beneficial. Clinical decisions are frequently made based on doctors' instinct and experience instead of the knowledge rich data hidden in the database. This type of working leads to undesirable blunders, errors and inordinate therapeutic costs which affects the quality of service given to patients. Some researcher suggested that integration of clinical decision support with knowledge-based system for patient disease diagnosis could lessen medical mistakes of diagnosing, enhance patient safety, reduce undesirable practice, and improve patient outcome. This suggestion is promising as data modelling and analysis tools, i.e. data mining, can possibly produce a knowledge-rich environment which can help to essentially improve the outcomes of clinical decisions. Data mining is the way towards finding already obscure examples and patterns in databases and utilizing that data to construct prescient models. In healthcare, data mining is winding up progressively. Healthcare industry today generates large amount of complex data about patients, hospitals resources, disease diagnosis, electronic patient records, medical devices, etc.

Revised Manuscript Received on September 21, 2020.

* Correspondence Author

Thejas Bharadwaj*, Computer Science Engineering, Vellore Institute of Technology, Vellore, India. Email: thejasbharadwaj98@gmail.com

Aatish Balaji, Computer Science Engineering, Vellore Institute of Technology, Vellore, India. Email: aatish2098@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

II.STATE OF THE ART

As per review of WHO, 17 million total worldwide deaths are because of heart attacks and strokes. The deaths because of heart disease in various countries happen due to work over-burden, mental pressure and numerous different issues. On the whole it was found as primary reason behind death in adults. Diagnosis is a complicated and important task that should be executed precisely and effectively. The diagnosis is generally made, based on doctor's experience and learning. This may lead to undesirable outcomes and over the top unaffordable medical costs of treatments. Therefore, a need for medical diagnosis system is raised that can take advantage of collection of data base and decision support system. This system can help in diagnosing disease with proper medical tests & effective treatments. The diagnosis of diseases is one of the most important and vital steps in medicine. Unfortunately, all doctors do not have a speciality in every field, also there are many places which have a shortage of certain kind of doctors due to which a lot of avoidable deaths occur. Therefore, an automated system which can combine all of their capabilities will be exceedingly helpful specially in those areas with shortage of medical help. Effective and accurate implementation of automated system needs a comparative study of various techniques available. Also, in [1] Hallick has proposed that CRM in healthcare can help promote disease education, prevention, and wellness services. Kolar [2] and Veletsos [3] have reported that Florida Hospital have been using data mining to segment Medicare patients as well as to develop applications that can enable credit scoring, debt collection, and analysis of financial data. Rafalski [4] has studied Sinai Health System's use of data mining for healthcare marketing and CRM. Not only that even pharmaceutical companies can benefit from healthcare CRM and data mining. Also, now it is possible to keep track of which physicians prescribe which drugs and for what purposes, pharmaceutical companies can decide whom to target, show what is the least expensive or most effective treatment plan for an ailment, help identify physicians whose practices are suited to specific clinical trials (for example, physicians who treat a large number of a specific group of patients), and map how an epidemic is moving through an area so that appropriate medical staff can be assigned. Pharmaceutical companies can also apply data mining to huge data sets to predict how a patient's genetic makeup might reach to a certain drug therapy. Data mining products such as SAS Enterprise Miner (SAS Institute, Inc., Cary, NC) are usually included in data mining application suites for certain application areas such as fraud and abuse detection, customer relationship management, and financial management. For example, Morgan Stanley uses SAS Enterprise Miner as part of its customer relationship management system. There are obvious advantages to an automated surveillance system, regardless of whether it is based on data mining or statistical methods such as statistical process control. These include (1) a reduction of time and effort on the part of the end user; (2) the ability to examine multiple areas simultaneously; (3) a decreased potential for human error; (4) data presented in the correct format; and (5) data that are accessible anytime and

anywhere. Automated surveillance systems raise citizen concern over privacy. But it is possible to extract useful and actionable information from records which has been stripped of private patient information. Military medical researchers in the United States are using systems that gathers data from military medical facilities worldwide as well as from other healthcare sources. The system detects outbreaks (e.g. the Norwalk virus in San Diego in 2002) in cases where the day to day doctors may not be able to see the big picture and it monitors progression of diseases (e.g. West Nile virus and listeriosis). There are also other factors boosting data mining's popularity. For instance, as a result of the Balanced Budget Act of 1997, the Centres for Medicare and Medicaid Services must implement a payment system based on classifying patients into mix groups, using experiential evidence. More details on data mining techniques can be found in Berry and Linoff [5] Healthcare Data Mining Applications. Generally, these data are grouped as the evaluation of treatment effectiveness; management of healthcare; customer relationship management; and detection of fraud and abuse. Data mining applications can be developed to evaluate the effectiveness of medical treatments. By comparing and contrasting causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective. Along this line, United HealthCare has mined its treatment record data to explore ways to cut costs and deliver better medicine [6]. It also has developed clinical profiles to give physicians information about their practice patterns and to compare these with those of other physicians and peer-reviewed industry standards. Similarly, data mining can help identify successful standardized treatments for specific diseases. In the Seton Medical Centre, data mining is used to decrease patient length-of-stay, avoid clinical complications, develop best practices, improve patient outcomes, and provide information to physicians—all to maintain and improve the quality of healthcare. According to a report by McKinsey, 50% of Americans have one or more chronic diseases, and 80% of American medical care fee is spent on chronic disease treatment. With the improvement of living standards, the incidence of chronic disease is increasing. The United States has spent an average of 2.7 trillion USD annually on chronic disease treatment. This amount comprises 18% of the entire annual GDP of the United States. The healthcare problem of chronic diseases is also very important in many other countries. For example, in china according to reports provided by the Chinese govt almost 87% of deaths are attributed to chronic diseases Therefore, it is essential to perform risk assessments for chronic diseases. With the growth in medical data [6], collecting electronic health records (EHR) is increasingly convenient [8]. Besides, [10] first presented a bioinspired high-performance heterogeneous vehicular telematics paradigm, such that the collection of mobile users' health related real-time big data can be achieved with the deployment of advanced heterogeneous vehicular networks.

In recent years, cloud computing gained a great attention in HCS applications due to its ability to provide different medical services over the internet. Cloud computing allows applications to provide infrastructure services to big numbers of stakeholders with assorted and dynamically changing requirements [7]. Technically, cloud is composed of databases, virtual machines server resources etc. Datacentres are containing a big number of resources and list of different applications. Cloud computing uses the virtualization technique which permits to share a single physical instance of a resource or an application among various stakeholders and enterprises [9]. It does this by allocating a logical name to a physical storage and providing a pointer to that physical resource when requested.

III. PROBLEM STATEMENT

Almost all major hospitals collect basic information about patients including billing information, and inventory. Although some hospitals are using predictive analysis using decision support system, they are very limited in number. The data gathered is not properly updated on a regular basis before doing the prediction which results in improper and inaccurate prediction on the data. Also, it is seen that the data is not properly pre-processed before doing the predictions on that data. The systems in present can answer simple queries like “Average age of patients who have heart disease?”, “How many surgeries had resulted in hospital stays longer than 10 days?”, “Identify the female patients who are single, above 30 years old, and who have been treated for cancer.” However, they cannot answer complex queries like “Identify the important preoperative predictors that increase the length of hospital stay”, “Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?”, and “Given patient records, predict the probability of patients getting a heart disease.” Clinical decisions are often made based on doctors’ intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. COVID-19 has led the world into a major crisis where the most strenuous efforts are being put in by the healthcare department. This kind of analysis will help the hospitals and doctors to determine the kind of support they require that is types of medicine in demand, the support machinery such as ECGs, ventilators etc. Once the trend is identified, the model can be used to predict the requirements of another place with a similar population density thereby helping in estimation of the required resources and being prepared beforehand for the upcoming events in the best way possible. This methodology can be ensued by various cities of India to minimize the catastrophic effects of this pandemic and also keeping them ready for any such reoccurrence as the data culminated over the years can be used to improvise the accuracy of the abovesaid model.

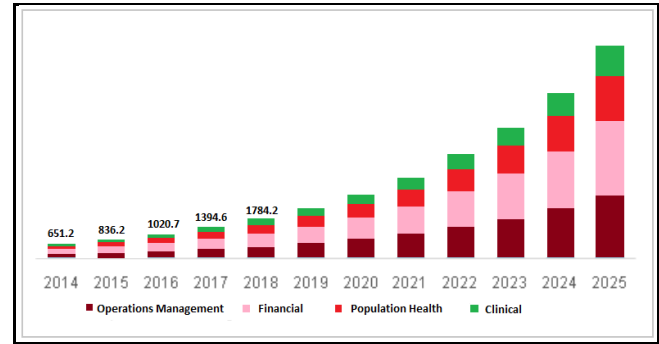


Fig. 1. Trend showing the demand of predictive analyzing healthcare system in the upcoming years

The Fig.1 depicts the high increase in demand for the predictive system from the year 2014 to year 2025 for the fields of operations Management, Financial, Population health and clinical. It can be seen from the trend that the demand of predictive health model is going to be in great demand in the upcoming years. Also, more advanced predictive systems are required as the world is likely to depend on the digital prediction in the healthcare system in by 2025.

IV. PROPOSED SYSTEM

The data of patients acquired from various hospitals is stored on the Customer Relation Management cloud which can be used for analysis like prediction of which are the common disease that are growing at a rapid rate, which type of disease occur for a particular category on basis of gender, age, etc. Our system aims at achieving three main limitation of the current data analytic system: 1) Lack of regular updating of Hospitals data on the servers 2) Improper data pre-processing before the analysis of the data 3) Irrelevant selection of model for doing the predictions. This system will try to regularly poll the data on a time span of 60 minutes. This will provide the system current data for analysis. This will decrease the human intervention on the server and will eliminate the manual updating of data. Also, a detailed data pre-processing will be carried out on data in order to assure that the missing data is properly filled and improper predictions are not done because of missing data. The pre-processing will be done with the help of techniques like binning, clustering, regression. Once the complete data is received after doing the pre-processing the data will be given to the analysing system which will analyse the data and provide the detailed analysis of the requested query. For the prediction four models namely random forest, neural network, linear regression, support vector machine will be used. The comparison will suggest which model is best for the given query. This is done because for different set of query different model are suitable. So, after passing the query through all the models will completely eliminate the risk of improper prediction. Thus, the model assures maximum guarantee about the predicted output. Moreover, the model will be useful for checking for the trends in disease with respect to different parameters like age, gender, economic status, location, etc.

The results achieved from both prediction and classification based on trends can be used by hospitals, government and pharmaceutical companies.

A. System Architecture

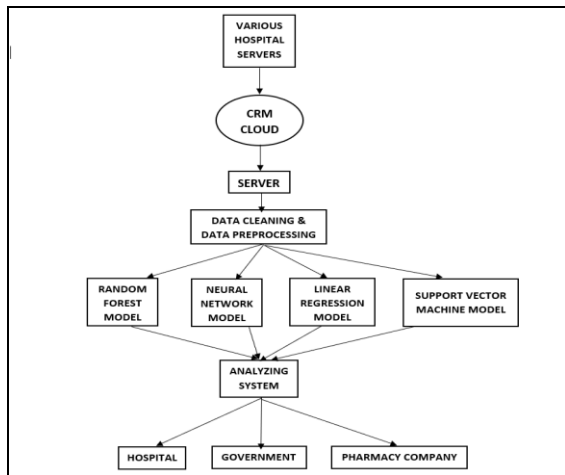


Fig. 2. System Architecture

Hospitals keep record of their patient’s details which is saved on their database, which contains various fields such as age, weight, allergies, symptoms etc. Several hospitals provide this data to a particular Customer Relationship Management (CRM) company like Salesforce which does analysis of the data provided to them to help the hospitals, government organisations and medicine companies. The data provided is polled after every certain amount of time and stored in a server. In this server, the data is put through different types of classification algorithms. The algorithm which is able to classify the data with the highest accuracy is used in predicting the missing points of the data provided by the hospitals and also in drawing conclusions which is helpful for the hospitals as it provides them with a clearer view about the diseases spreading in the city and employ such specialist doctors. The government can raise awareness about particular things which are the cause for the rapid increase of that disease whereas medicine companies can increase the amount of drug being produced as per the requirement highlighted by the analysis of data.

B. External Architecture

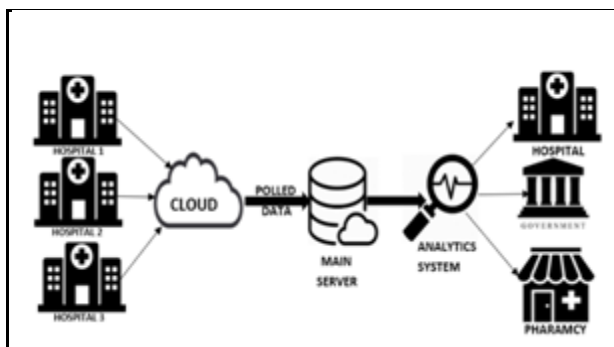


Fig. 3. External Architecture

In this figure, various hospitals provide data to the cloud storage of the CRM company from where the data is polled into the main server where analytics is performed and conclusions are drawn which are then forwarded to the

government, hospitals, medical companies and whosoever contacts the CRM company for the analysis and conclusions.

C. Methodology

1. Polling



Fig. 4. Polling Scenario

The Customer Relationship Management (CRM) which is responsible for doing the prediction and analytics stores the information in the cloud. The cloud is accessed by an authenticated Software which polls the data. The software (like MuleSoft) is designed through a script written in XML. The configuration is done in such a way that the software requests data from the cloud servers after every 60 minutes. Whenever such a request is made the cloud transfers the data from the cloud to the server. It is to be noted that only the data which has been added to the server in the past 60 minutes is only put into the local server, hence making the data redundant. Once the information is present in MuleSoft, a file of the data is generated which has the complete dataset which has been collected since the beginning to the last time data was polled in. This make the dataset dynamic and keeps improving the prediction model as well as improving the analytics.

2. Random Forest

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. What it does as an overview is that it creates a multitude of decision trees which are all based on the sample data and then gets the prediction output from all of them. After that it select the best solution out of all of them through means of voting It is an ensemble method which is better than other single decision trees because it reduces the overfitting by averaging all the results. Random forests address one of the primary problems of decision trees, which is overfitting to the preparation information. Since decision trees depend on an arrangement of tests on traits, a decision dependent on a terrible guideline is probably going to have a bad effect on tests that come after it, and on the aftereffects of the model. Random forests maintain a strategic distance from this by structuring many separate trees, each with various branch structures, and taking an accord of their outcomes.

3. Neural Networks

Neural systems are a regularly utilized and understood AI/data mining procedure.

They have numerous hubs/neurons, every one of which is associated with various different hubs. Hubs are masterminded into (at least three) layers, the info layer, shrouded layer(s) and yield layer. Every hub in a layer is associated with the majority of the hubs in the promptly neighbouring layers. For relapse, the yield layer may comprise of a solitary hub that, having gotten contributions from the hubs in the past layer, will give a solitary incentive as yield. All together for a neural system to be valuable, it must be trained. This is done by giving more input values. These trained values are put through the nodes in each layer to the nodes in the resulting layer. At the point when a node gets a lot of data sources, it will apply "weights" to every one of them. It will likewise apply a capacity to the weighted data sources, and pass on the yield to every one of the nodes it is associated with in the following layer.

4. Support Vector Machine

Support Vector Machine (SVM) is a supervised AI calculation which can be utilized for both characterization or relapse difficulties. Be that as it may, it is for the most part utilized in arrangement issues. In this calculation, we plot every data as a point in n-dimensional space (where n is number of highlights) with the estimation of each component being the estimation of a specific coordinate. At that point, we perform characterization by finding the hyper-plane that separate the two classes.

5. Linear Regression

Linear Regression is an AI algorithm dependent on regulated learning. It plays out a relapse task. Regression model is an objective prediction scheme dependent on independent factors. It is generally utilized for discovering the connection among factors and estimating. Various regression models differ depending on – the sort of relation among dependent and autonomous factors, and the number of independent variables used. Internal Architecture

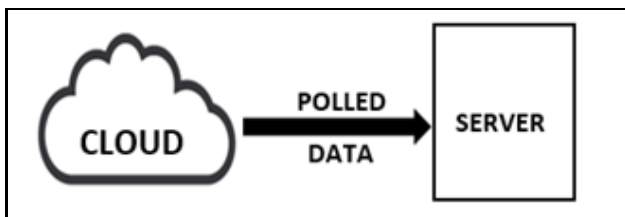


Fig.4.1 Internal Architecture-1

In this figure, the data stored in the cloud storage (a model of cloud computing in which data is stored on the internet) is polled. Polling is the repeated transfer of data from cloud to the local server, the data centre where computation on the dataset can be done, through the use of a software like MuleSoft. In polling, the new data which has arrived in the cloud is transferred to the server. In the other words it can be said that incoming data is appended to the existing data. The server is connected to the cloud storage via the internet and the new data arriving at the cloud is polled after a certain period of time. The time period of the data to be polled is set by the CRM company. The time period is set in such a way that data is pushed onto the local servers only when there is a significant addition into the cloud storage and the new data will bring about changes in the analysis of the data.

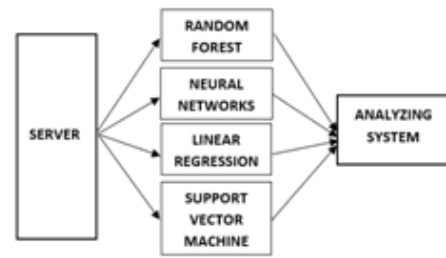


Fig.4.2 Internal Architecture-2

The data which has arrived at the local storage after being polled in from the cloud is then classified. Four classifiers namely Random Forest classifier, Neural Networks, Supported Vector Machines and Linear Regression Model are applied. The model which is best able to predict the missing data that is with the highest accuracy is chosen to analyse the system. In the Random forest which is a supervised learning algorithm an assembly of decision trees are constructed when training and then outputting the prediction. A support vector machine is a model that stores information sections as focuses in a n-dimensional space (n = the quantity of highlights in the informational collection). It is a model preparing information, and the majority of the information passages must be categorized as one of two classifications. In view of these information passages, a support vector machine tries to discover a hyperplane (basically a line mapped to the n-dimensional space) that will ideally isolate the data points having a place toward various classifications. Linear regression is a very simple type of model; it seeks to find a straight line of best fit for the data. It will take in all of the data features as input and find the equation of the line that best maps these to the correct output values. In order to find this line, linear regression methods often use the method of least squares. This means that the model will seek to minimise the sum of the squared errors of each data point from the line.

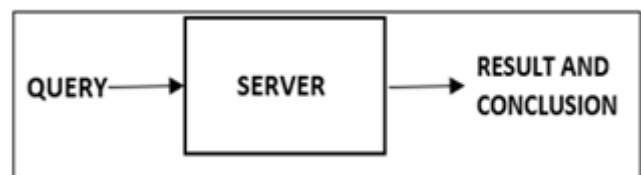


Fig. 4.3. Internal Architecture-3

Queries are made to the server which can be answered through proper analysis of the information stored on the database server. The meaningful analysis made from the data can be queried and forwarded to the concerned authorities so that they can make amends to their business models accordingly.

V.DISCUSSION OF ANALYSIS

The system comprises of a synergy of data mining, machine learning, data polling, cloud computing, data analysis and data prediction. Data Mining techniques provide us an efficient and high-quality data analysis as well data prediction based on our requirements.

These predictions and analysis serve as a boon for the health care systems as they can now predict the disease which are likely to occur in future as well as analyse the current trend in healthcare data. Consider the scenario in which the polled data from the mule software is obtained. This data is used for training the data mining model. This model can further be proved useful for the various healthcare predictions. For simplicity we are considering a small training data set consisting of attributes of age, sex, weight (in lbs), height in inch, IQ, unit of alcohol consumption per day, number of cigarettes per day, active/passive, health score. With the help of these data we will train our model and try to predict the values with the help of various data mining models and will compare each and every result and finally conclude the best from all results.

Table 1: Training Set

Age	Sex	Weight in lbs	Height in Inch	IQ	Units of alcohol per day	Cigarettes per day	Active	Health Score (high is good)
32	Female	101	67	110	0	0	Inactive	251
47	Female	138	67	84	2	0	Active	193
25	Male	166	73	106	2	5	Inactive	167
75	Male	184	73	99	2	0	Inactive	93
72	Female	105	65	99	2	0	Active	150
59	Female	133	62	107	0	0	Inactive	120
31	Female	162	65	95	1	0	Inactive	167
72	Male	167	70	94	1	0	Inactive	90
53	Female	152	66	99	0	0	Active	181

Now for this scenario we are considering a small training data set of 9 records. For greater accuracy dataset with more records must be used for training. After successfully training our model, we will now try to predict the following five records health score.

Table 2: Prediction Set

ID	Age	Sex	Weight in lbs	Height in Inch	IQ	Units of alcohol per day	Cigarettes per day	Active	Health Score (high is good)
1	65	Male	191	69	104	0	0	Active	??
2	46	Female	99	66	103	0	0	Active	??
3	51	Female	135	71	105	1	0	Inactive	??
4	37	Female	126	66	96	1	0	Active	??
5	77	Female	114	67	99	0	0	Active	??

After giving the input to our model a detailed result of all the values from various used models will be given as output in which the average of all the values will be chosen as the final output answer. This will remove the chances of false prediction and will help to get the best possible prediction.

Table 3: Result

Person ID	Random forest model	Neural Network Model	Linear regression model	Support Vector Machine Model	Final Answer
1	125.2	122.1503	170.98028	123.352	122.9
2	238.8	241.6606	235.13031	226.014	235.4
3	149.5	134.5609	160.24534	144.72	147.3
4	244.1	257.3349	238.95514	258.332	249.7
5	153.9	161.3498	157.54189	154.027	156.7

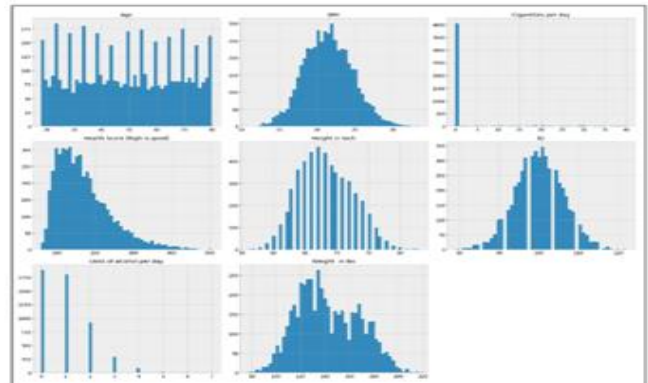


Fig 5: Histogram Distribution on various parameters

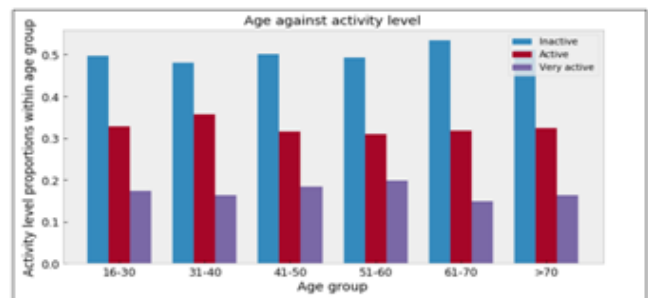


Fig 6: Classification on various age groups

After the data prediction is done, we can use machine learning technique to analyse the data based on the various parameters like age, gender, height, weight, etc. This analysis can be helpful for hospitals, pharmacy companies, government. The analyses show various histogram distribution based on various parameters. Classification based on the various age groups for the parameters such as the inactive, active, and very active on the dataset.

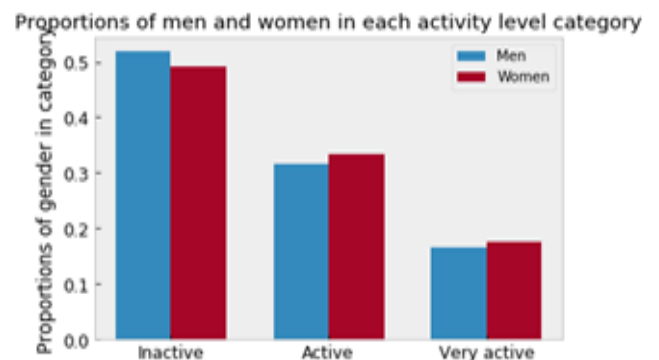


Fig 7: Analysis based on gender for different parameters

Analysis based on the gender for the parameters of active, inactive and very active population. Thus, these analyses will be helpful for the government for spreading awareness if some disease is likely to happen. Also, the hospital or medical college can provide admission to students based on the requirement. Also, the pharmacy companies can produce medicine as per requirement.

VI.FUTURE SCOPE

The system implemented is used in healthcare system whereas the same arrangement can be modified and be applied to different sectors such as automobile industries who can, through the help of a Customer Relationship Company (CRM), draw conclusions about their product and work according to the reports they get. Similarly, all product-based companies can benefit from this scheme. The scheme itself can be enhanced to work better by using more classification algorithm to help improve the prediction model therefore improving the data analytics system too. The analysing model works according to the queries made by the companies which can be improvised through the use of Artificial Intelligence where the model will itself know, through the data provided, what kind of analysis is to be done. This will make the entire system self-sustaining where information is fed into it from one end and from the other end conclusions are provided. Another area where this kind of polling can prove useful can be seen right now as of writing this paper. The world is in chaos because of the pandemic (COVID-19). Although not perfect this system could have been used to predict the exponential increase in the number of cases and help in identifying in what places is there a need for more personnel. Yes, the world did catch on to the growing number of cases but that is because of the huge scale of the spread. If it was concentrated to a small area with limited capabilities in terms of health care it would have been difficult to track the spread and extent of the disease. This is where our system could come in where in the system would have been able to alert the officials of a possible outbreak and they would be able to react much faster as compared to using traditional ways to detect an outbreak.

VII.CONCLUSION

Healthcare has an ocean of data whose analysis if done properly can be useful to draw conclusion which would be helpful in many sectors of life. Since data is being polled constantly into our analysis and prediction system it is increasingly being improved with each polling process. Unlike other mining and analysis system, data is continuously being which has a huge advantage over other systems in this way. Also, we are implementing four different algorithms, namely, Random forest model, Neural Network, Support Vector Machine (SVM) and Linear Regression Model, to classify the information obtained from the servers. The algorithm which is best able to classify the data is used to predict the missing information Analysis of the data gives a clear image about the condition of the healthcare system prevailing and how it can be improvised by the companies, organisations and hospitals who obtain the analysed data. The hospitals know which medical

personnel have a higher demand, government organisations can know what kind of disease is spreading rapidly and how it can be contained whereas the medical companies will know which medicine has a higher demand and make changes to their production plans accordingly.

REFERENCES

- Hallick, J.N. (2001), Analytics and the data warehouse. Health Management Technology.
- Kolar, H.R. (2001). Caring for healthcare. Health Management Technology,
- Veletsos, A. (2003). Getting to the bottom of hospital finances. Health Management Technology.
- Rafalski, E. (2002). Using data mining and data repository methods to identify marketing opportunities in healthcare. Journal of Consumer Marketing.
- Berry, M.J.A. & Linoff, G.S. (1997). Data Mining Techniques: For Marketing, Sales and Customer Support. New York: John Wiley & Sons Inc.
- Palaniappan, S., & Awang, R. (2008, March). Intelligent heart disease prediction system using data mining techniques. In 2008 IEEE/ACS international conference on computer systems and applications (pp. 108-115). IEEE.
- Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-48.
- Bhatla, N., & Jyoti, K. (2012). An analysis of heart disease prediction using different data mining techniques. International Journal of Engineering, 1(8), 1-4.
- Perez-Iratxeta, C., Bork, P., & Andrade, M. A. (2002). Association of genes to genetically inherited diseases using data mining. Nature genetics, 31(3), 316.
- Kharya, S. (2012). Using data mining techniques for diagnosis and prognosis of cancer disease. arXiv preprint arXiv:1205.1923.
- M.A.Saleem Durai, N.Ch.S.N.Iyengar, —Effective Analysis and Diagnosis of Lung Cancer Using Fuzzy Rules; International Journal of Engineering Science and Technology, Vol. 2(6), 2010.
- Asha Rajkumar, G. Sophia Reena, —Diagnosis of Heart Disease Using Datamining Algorithm; Global Journal of Computer Science and Technology, Page | 38 Vol. 10 Issue 10 Ver. 1.0 September, 2010.
- Shantakumar B.Patil, Y.S.Kumaraswamy, —Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network; European Journal of Scientific Research, ISSN 1450-216X Vol.31 No.4, 2009.
- Rupa G. Mehta, Dipti P. Rana, Mukesh A. Zaveri, —A Novel Fuzzy Based Classification for Data Mining using Fuzzy Discretization; World Congress on Computer Science and Information Engineering, 2009.
- Sellappan Palaniappan, Rafiah Awang, —Intelligent Heart Disease Prediction System Using Data Mining Techniques; 978-1-4244-1968-5/08/\$25.00©2008 IEEE.

AUTHORS PROFILE



Thejas Bharadwaj, is currently pursuing Btech in the field of Computer Science and Engineering at Vellore Institute of Technology, Vellore. During the course of his engineering he has completed several internships in the field of Web Development and Software Development. He has shown his interest in the field of Machine learning through his multiple online course completions on platforms like coursera and Udemy. His area of interest includes Machine Learning, Big data and Artificial Intelligence.



Aatish Balaji, is currently pursuing Btech in the field of Computer Science and Engineering at Vellore Institute of Technology, Vellore. During the course of his engineering he has completed several internships in the field of Web Development and Data science. He has shown his interest in the field of Machine learning through his multiple online course completions on

platforms like coursera and Udemy. His area of interest includes Machine Learning, Big data and Artificial Intelligence.