VOL. 13, NO. 14, JULY 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A FRAMEWORK FOR DECISION MAKING AND QUALITY IMPROVEMENT BY DATA AGGREGATION TECHNIQUES ON PRIVATE HOSPITALS DATA

Syed Ahmed Yasin and P. V. R. D. Prasada Rao
Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India
E-Mail: syedahmed.cse@gmail.com

ABSTRACT

Predictive Analytics using Big Data is emerging field which help to make quick, accurate decisions from structured as well as unstructured data. There are vast fields like Healthcare, Education, and Weather Forecasting where Predictive Data Analytics can help us for getting insights of data. The Biggest challenge for Healthcare Industry is huge data, rapid generation of speed & complexity of data. In this paper, we have concentrated on how Data Science& Big data analytics can help us for improvement in Health Care Analysis and Prediction of Accurate results. Tools used in Data Science and Big data are also discussed. For experiment we have used two Data Mining Techniques; Decision Tree and Naïve Bays which are applied on UCI Data set and Actual Hospital Data set. Performance Analysis is done for 13 and 15 medical attributes of data.

Keywords: Data science, big data, healthcare, data analytics, predictive analytics.

1. INTRODUCTION

Data Science is field of similar to Data Mining where scientific methods, algorithm, and process are applied on data to extract useful information, insights from data in semi structured, unstructured, and structured format. Data science field is combination of machine learning, statistics and data analytics for analysing and understanding the real phenomena with data. There is strong relationship between data science and healthcare. Developing medical application with the use of big data & machine learning techniques has become more necessary within previous years; in fact most of the organization from totally different Sector relies more and more on data extracted from vast volume of data. Big data has potential for supporting a range of healthcare or medical function such as surveillance for disease, population insight, and medical decision support system. Many frameworks are designed and promoted for health care solutions which are based on disease focus and patient focus in which active participation of doctors as well as patient are required. As medical sector data complexity is high they show strong relationship among them, Hence tasks such as data acquisition, data simplification, data complexity, relation between an attribute, selection of target attribute for medical data analytics requires domain specific tools and techniques [1]. Due to rapid advancement in HER, e-Health, m-Health, and smart health led to design and development of health care decision support framework for support of accurate diagnosis, medicine and patient care.

The Latest research shows that the comprehensive healthcare solution has been proposed with various layered services In this paper we have presented different analytic techniques that exist in patient focus or centric healthcare system where applicability to all we be applied. In addition the implication of data mining, big data analytics tool in development of healthcare system is also presented, detail survey is undertaken. Big data is trending term that describe big amount of unstructured, structured and semi structured amount of data that can be mined for useful information. Big Data comprise information with having huge volume which is past the capacity of usually utilized conventional tools, to catch, process, and oversee information inside a particular time. As big data not refer to any particular quantity, so it can often use when we are dealing with PB and EB of data. When dealing with such huge measure of information enterprises and other association confront ready to make, control, troubles in having the capacity to make, control, and mange enormous information, huge information in genuine its concern in business. Big data term was coin by John mashey in the 1990's. Big data consist of structured, semi structured, & unstructured data but main area of focus on unstructured data processing and mining. Big data requires computerizing technique, algorithms which must have new method of integration to find out knowledge from complex and big scale data. Big data growth can be defined by 3D model i.e. amount of data (Volume), Speed of in or out data (Velocity) and range of data type or structures (Variety). The 3 Vs have been extended to other supplementary characteristics such as variability and veracity which refer to inconsistency and quality of data respectively [2].

All 5 V can be taken as complexity measures for big data. For knowledge extraction from complex data we required data analytics involving data science and machine learning. Data availability is everywhere, there is rapid rate of data growth, and for these issues big data analytics can play a major role5 Vs of big data are is as follows: Volume, Velocity, Variety, Variability, &Verracity.

Table-1. Indicates 5 vs of big data.

Characteristics	Description
Volume	Huge Amount Of Data created by sources
Velocity	Speed at which data is collected, analysed, and shared
Variety	Diverse sources of data
Variability	Refers to change in data rate, data format & data semantics
Veracity	Indicate the quality of data produced

Below Diagram shows how big data is beneficial for cost reduction, fast, better decision making and for new product &services, all these factors can be vital for health care improvement.

1.1 Applications of big data

1. Government: All parts of government body such as local and central are required to be collaboration for data analysis, so that efficient results will be obtained in format of productivity, and cost. Worldwide it's observed that U`S and UK has done good use of big data for government formation and taking public surveys.

2. Finance: Every customer who uses banking and finance service generates electronic records that must be retained as per bank policies with the help of big data bank finance services not storing data but also using it to generate business insight and add value to it such as fraud detection. By applying machine learning technique they are able to define normal activity based on customer history and differentiate it from unusual behaviour indicating fraud. The data analysis system shows and suggests immediate preventative actions such as blocking of transaction, which stops fraud activity before it occurs and improve profitability. Other finance application include customer segmentation, personalized marketing and risk management.

3. Manufacturing: Predictive manufacturing can be done with the help of data acquisition. Generated big act as input to predictive tools, highest level of benefits and profit can be achieved when big data analytics is used in manufacturing smartly.

4. Health care: Due to M-health, E-health and wearable technologies in health care volume of data is increasingly continuously. This include imaging data, patient generated data, sensor data, while extensive information in healthcare is now electronic ,its fit under the big data umbrella as most is unstructured and difficult to use. Big data analytics can help health care to improve by Providing personalized medicine, clinical risk intervention and predictive analytics, internal, external reporting of patient data.

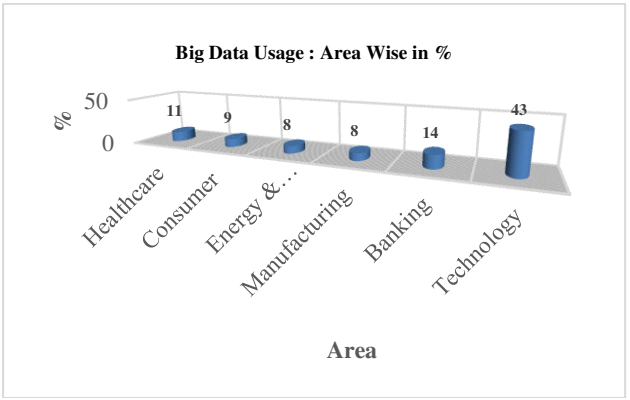
5. Education: With the help of learning analytics academic institutions can better understand student needs and proactively address them. Education sector can be benefited from big data such as competition for admission, student performance management, & indicator of student success.

6. Media: Traditional approaches of using specific media environment such as TV, newspapers, magazines as replaced by technologies which can reach targeted people at optimal time in optimal location. Other uses of media is targeted marketing, Data Capture.

7. Internet of thing: IOT and big data can work in conjunction. Data extracted from hardware devices such as media, companies, industry, and government to more accurately target their audience and increase media.

8. Information technology: IT departments can easily predict potential risks and issues before the problem occur. This can be possible when IT is in use with conjunction with machine learning and deep computing.

9. Sports: Big data can be used for training and understanding of competitor using sensor. Winner of particular match can be predicated by big data analytics. Future Performance of players can be predicated by big data analytics. Following graph shows area wise big data usage in percentage.



Graph-1. Area wise big data uses.

1.2 The impact of data science in healthcare system

In medical sector, the data science has a vital role for predictive analytics techniques, and machine learning as effective implementation of personalized medicine and treatment plans. In addition to this impact may be in telemedicine, social media, wearable sensor devices, also patient history clinical and diagnostic trials data, effectiveness of drug index. A typical patient healthcare system with involvement of stakeholders and other use case is described in Figure-1. When these all stakeholders work in collaboration and exchange data effectively, a feasible solution would be offered with patient focused and in cost effective manner. Considering importance of all stake holders in building a medical DSS, the following section gives an insight on their involvement for effective utilization of big data and data science sources.

Figure-1. Impact of data science in healthcare system.

1.3 Stakeholders of medical decision support system

Stakeholders are the actors which are directly or indirectly benefited and involved in progression of system, we have identified few of them. Description of all is as follows,

Patient: One of the end user of this framework is patient, who expects quality service at low moderate cost. Patient has opportunity to gain extra medical knowledge through social networks, clinical forums, these Sources helps patient to connect with similar person for gaining information such as symptoms ,side effect, drug information, review results.

Doctors: Due to many type of data sources such as clinical notes, lab results, imaging data and sensor data, such data can be grouped and stored in repository, which will improve public health, and faster response through effective analysis of diseases pattern.

Pharmacy researchers: The impact of data science reflects healthcare information in pharmacy and clinical research, which will help to build predictive model for decision making process and for building predictive model from it.

Insurance company: Insurance is key aspect of medical sector, where predication analysis technique can be applied based on frequency occurring diseases. Based on location, with minimum cost. Claim history data can be provided for important information for predication of pattern of authentic claims and outliers if any.

Service providers and operators: Medical operator highly relay on outcome of data source such as post treatment data generate from, phone calling for follow up, communication through mail, SMS, can also improve quality of data and result. Following Figures-2 demonstrate the all stakeholders.

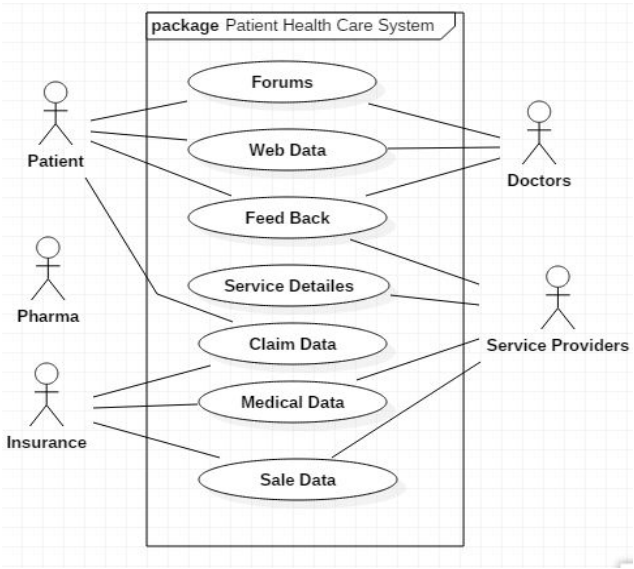


Figure-2. Stakeholders of patient health care system.

2. LITERATURE SURVEY

Ruogu Fa, *et al* has worked on computational health informatics. The goal was to study challenges and current big data mining techniques and directions for future work. Data pre-processing, data capturing is also discussed [1]. Junfei Qiu, *et al* explored deep learning, parallel and distributed learning, Active learning in their work Data Meaning & Pattern Training issued was discovered [2]. Jianguo Chen, *et al* worked on PTTP algorithm for big data and the Apache Spark Cloud Storage. A random forest improvement formula is performed for the PTTP model. The queue waiting time of every treatment task is expected supported the trained PTTP model. Intensive experiments and application results show that our PTTP formula and HQR system attain high preciseness and performance. Progressive PTTP formula supported streaming knowledge and a additional convenient recommendation with decreased path-awareness area unit prompt for future work. [3]. Javier Andreu Perez *et al* proposes huge information for health framework for medicine and health informatics. Focous was however these information science can have the benefit of Associate in Nursing integrated approach of piecing along totally different aspects of customized info from a various vary of information sources, each structured and unstructured, covering genetics, proteomics, metabolomics, additionally as imaging, clinical diagnosing, Associate in Nursing long continuous physiological sensing of an individual [4]. Asif Adi *et al* worked on Multi Disease Analysis for Improvement in Healthcare by use of big data, they opted for Medical Body Area Networks, MBANs change an eternal observation of patient's condition by sensing and sending recorded measurements like pulse rate, graphical record (ECG), blood heat, rate of respiration, chest sounds, and pressure use of Hadoop is finished for unstructured knowledge [5]. Nina S.etl worked on green healthcare with the help of big data, such as utilization of Electronic Medical Records (EMRs), (2) Telemedicine, (3) virtualization of servers utilized as a part of social insurance area, (4) virtualization of work areas utilized as a part of medicinal services related workplaces and (5) virtual joint effort. Security of patient protection, IT foundation electric vitality manageability, information assurance, and the proceeded with change in distributed computing for IT cost lessening alongside enhanced information insurance can be gotten with these structure [6]. Reena Duggal *et al* propose how to utilization of Big Data investigation procedure Map Reduce alongside Fuzzy rationale to coordinate gigantic measure of patient records with persistent coordinating [7]. Marco Viceconti et al proposes however huge information analytics may be with success combined with VPH technologies to provide sturdy and effective in silico medication solutions. So as to try to to this, huge information technologies should be additional developed to address some specific necessities that emerge from this application [8]. Xue-wen Chen, *et al* works on big data and deep learning challenges and issues. They trust that examination challenges postured by Big Data are convenient, as well as bring plentiful open doors

for profound learning. Together, they will provide major advances in science, medicine, and Business [9]. Maryam M Najafabadi *et al* give a concise review of profound learning, and feature ebb and flow look into endeavours and the difficulties to enormous information investigation [10]. Chun‐Wei Tsai *et al* discussed new issues for information examination, for example, protection, security, stockpiling, adaptation to non-critical failure, and nature of information [11]. Sara Landsat *et al* have completed a study of open source devices for machine learning with enormous information in the Hadoop biological community [12]. Jing Lu, Malcolm Keech *et al* paper provides a short review of however rising information technologies are often wont to give further price from obtainable health data like electronic patient records. The pertinence of NoSQL and large information technologies is taken into account with a stress on information storage and mining approaches [13]. Hiba Asri *et al* papers demonstrate challenges and opportunities in big data such as the genuine test in human services frameworks is the means by which to discover, gather, break down and oversee information to make individuals' lives more advantageous and simpler, by causative not exclusively to know new illnesses and treatments however conjointly to anticipate results at prior stages and construct period choices [14]. Seth Early proposes a framework for big data and predictive analytics which can be benefited in improvement in mining techniques and increment in marketing effectiveness [15]. Harsh Kupwade Patil *et al* proposes a framework for giant information security and privacy problems in health care. The main focus of the paper is on from reactive to proactive attention may end up in associate overall decrease in attention prices and eventually result in economic process [16]. Prof. Jigna Ashish, Dr. Priyanka Sharma etl papers is based on Big Data for Better Health Planning which shows how wearable monitors and other tools can be used for better health planning [17]. Bijesh Dhyani, *et al* works on Big Data Analytics using Hadoop Framework by using map reduce [18]. Ashwin Belle, Raghu ram *et al* papers use image and signal processing for genomics in medical analytics is done [19]. This survey paper portrays the significant advancements in human services alongside exchange about prevalent procedures, devices and database, along with details of single and compound data mining techniques used. In recent years, the main motto behind this paper is to study various algorithmic techniques used for predication of heart disease and compare the techniques to find out best techniques for predication and health care improvement

Mathematical Modeling:

Bayes rule: According to Bayes' theorem, mathematically the probability that is to be compute, $P(H \mid X)$ can be expressed in terms of probabilities $P(H)$, $P(X \mid H)$, and $P(X)$ as

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)}$$

Where, *H*=Hypothesis, P *(H)* =independent probability of h (prior probability), P*(X)* =independent probability of H, *P (X|H)* = conditional probability of *X* given H

Gaussian distribution with a mean μ and standard deviation defined by

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp - \frac{(x-\mu)^2}{2\sigma^2}$$

Where, $\mu = Mean$, $\sigma = Std\ Deviation$

Decision tree: Entropy(S) is a live of the quantity of uncertainty within the (data) set S (i.e. entropy characterizes the (data) set S).

$$H(S) = -\sum_{x \in X} p(x) \log^2 p(x)$$

Where, S = current (data) set that entropy is being calculated, p(x) =Proportion of the amount of components at x, to the amount of components in set S once H(S)=0, the set S is utterly classified (i.e. all components in S square measure of identical class).

Information gain:

Information gain IG (A) is that the live of the distinction in entropy before to once the set Sis split on Associate in Nursing attribute A. In alternative words, what proportion uncertainty in S was reduced once cacophonic set Son attribute A.

$$IG(A) = H(S) - \sum_{t \in T} p(t)H(t)$$

Where, H(S) = Entropy of set S, T= The subsets created from cacophonic set S by attribute A such, p (t) = The proportion of the amount of parts in t to the amount of parts in set S, H (t) = Entropy of set .Here the proof could be a scaling issue dependent solely on that's constant if the values of the feature variables area unit illustrious.

3. DATA SCIENCE AND BIG DATA TOOLS FOR MEDICAL DECISION SUPPORT FRAMEWORK

Many big data framework are designed towards meeting a specific healthcare objectives such as data collection, data pre-processing, data analysis, data interpretation and visualization .following section we focus on use of various big data tools that play a vital role in performing the task such as searching, indexing, integration, data processing, and visualization of data.

3.1 Machine learning tool

Health care sector requires machine learning tool to analyse the data and transform it to actionable knowledge by performing predicative analytics in support to intelligent and effective clinical activities.

Sky tree: Is a general purpose machine learning tool that uses AI to generate algorithmic techniques for performing advance analytics technique. This tool can process massive amount of data with high accuracy .This tool is used for similarity searching, market segmentation, clustering, predictive analytics, outlier detection, & development of recommendation system.

Karma sphere: It creates big data platforms that mines and do analysis of web data, cell phone, and social media data by using Hadoop. It support GUI features for data navigation to predict and analyse the pattern.

Big ML: It provides multiple machine learning platforms such as classification, clustering, association ,it integrate the features of machine learning with cloud environment for development of system which will be highly scalable, flexible, reliable and cost effective.

Apache mahout: Its open source library of machine learning used in Hadoop environment for execution of machine learning techniques in distributed environment. This tool has support for techniques such as, Clustering, Pattern matching, Classification.

3.2 ETL tool

As data is generating continuously and volume and velocity is increasing, there is demand for data integration tool for data aggregation from disparate sources.

Pentaho: This tool provides end to end data integration to support user for analysis of data from heterogeneous sources of data. It also provides user interface for creation of visual information flow to perform data Transformation and Integration.

Palantir: Its ETL tool which combines the data from different data sources, it also enables analytical techniques for development of expert model for tracking sequence of procedures and clinical data to manage healthcare diagnosis.

Jitter bit: It offers a platform for medical care system to access clinical data by grouping data with unstructured to structure in multiple standard formats.

Informatica: It offers a wide range of data management services that include analytics, integration and governance of health care data for improvement in healthcare.

Attunity: This tool is used for data integration from data sources such data warehouse, Hadoop in faster manner with less coding.

3.3 Searching and processing tool

As medical data is present in all formats of structured and structured text there is strong need for searching and indexing tool for performing optimized full searching on clinical data.

Cloudera impala: It's useful for apache Hadoop file format which offers low latency, high performance SQL queries.

Apache lucena: Apache Lucena: It's a high performance compartmentalization system that provides powerful and correct full text search facility for kind of applications across completely different platforms.

Google drenel: This tool uses multi-level execution trees for query processing. It uses multi-level execution trees for query processing.

3.4 Real time and stream data processing tool

Current IOT based sensors devices found in healthcare sector promotes the data processing from heterogeneous data sources to be carried in real time manner.

Hana: It's an in memory analytics platform which supports database services, app development, data adminsrstration, advance analytics, & Data openness.

Apache Kafka: Its distributed streaming platform which is used for development real time data processing and applications.

Splunk: It is used for real time application where knowledge is to be extracted from heterogeneous sources of data.

SQL stream blaze: This tool captures data from all sources in all formats at various velocities.

Apache strom: It's used in real time any tics such, interactive operation system, online machine learning expert system, and ETL.

3.5 Visual analytics tool

Visualization of data help to identify trends, pattern that may include outliers, clusters, association discovery and time series analysis for improvement in clinical healthcare delivery and public health.

Tableau: This tool has ability to transform huge amount of complex data sets into pictures .It combines advances in database and computer graphics to analyse big data with less resources.

Qlik: Enables health care organization to analyse hospital, financial.

Jasper soft: This tool support effective decision making with the help of interactive dashboards, analytics, and reports. It provides visualization for Mongo DB, Cassandra etc.

Following is summary of tools:

Table-2. Summary of tools.

Data Analytics Task	Tools Used
Machine Learning	Sky tree, Karma sphere, Big ML
ETL	Pentaho, Apache Mahout,
Searching & Processing	Splunk, SQL Stream Blaze, Apache Strom
Real Time & Stream Data	Hana , Apache Kafka,
Visuals Analytics Tool	Tableau, Qlik, Jasper soft

4. ISSUES OF BIG DATA

Following section describe the challenging issues of big data analytics

Data growth: Dealing with big data is big challenge as data is to be stored for analysis of information. In digital universe reports IDC estimates that quantity of data holds on in world. IT system is doubling concerning each year. By 2020 the mixture total are spare to fill a pile of tablets that ranges from earth to moon six.6 times Most of the information is in unstructured format, meaning it does not resides in database. Records, photographs, sounds, recordings and other unstructured information can be hard to seek and break down. Organization are dealing data growth with different technologies, converged infrastructure, Software outlined storage will build it easier for corporations to scale their hardware and different technologies like compression, duplication, and exhausting will cut back the quantity of area and value related to huge knowledge storage.

The solution of above data growth problem can be solved by tools like, NO SQL database, Hadoop, Sparse and big data analytics software, machine learning can help industries to discover knowledge from huge amount of data.

Timely generation of insight from data: Most Common goals associated with big data project include decreasing expenses through operations cost efficient, establishing information driven culture, propelling new item and administration offerings.

To achieve the speed some organization are looking to new generations of ETL and analytic tool that reduces time it takes to generative report. Software with real time analytic can be performed for insight from data.

Data governance: Industries area unit getting similar information from completely different system the method of obtaining such records to agree, additionally as creating specified records area unit correct, usable, and secure, this method is named as information governance.

Data integration: Data variety of big data leads to challenge in data integration, sources of data can be enterprises, social media, email system, and combination of all can be done to make reports. Organization are exploitation form of ETL and information integration tool to form method easier, however several organization downside of information integration isn't solved

Security of big data: Big data stores can be attacked by hackers, organization use security measures such as access control, data encryption and data segregation Issues generally arises with data privacy and authentication

Organizational resistance: Not only technological aspect but also people management can be issue in big data. Following are the issues of organization:

a) Insufficient Organization Alignment

b) Lack of middle management alignment & understanding

5. PROPOSED SYSTEM ARCHITECTURE AND RESEARCH OBJECTIVES

1. To evaluate the prediction capability of various machine learning algorithms applied on medical datasets.
2. To study the criteria used by the patients in choosing the hospital service providers & to examine the level of satisfaction of patients and the factors influencing their level of satisfaction.
3. To evaluate the quality of health care services provided by the private hospitals & to offer suggestions in enhancing the quality of hospital services and patient satisfaction.

Figure-3, represents key elements of proposed system, it consist of health data, price data, and actors of the system. For data analytics and knowledge discovery we have collected two types of data one is price data of individual services provided and health data which is required for data analysis.

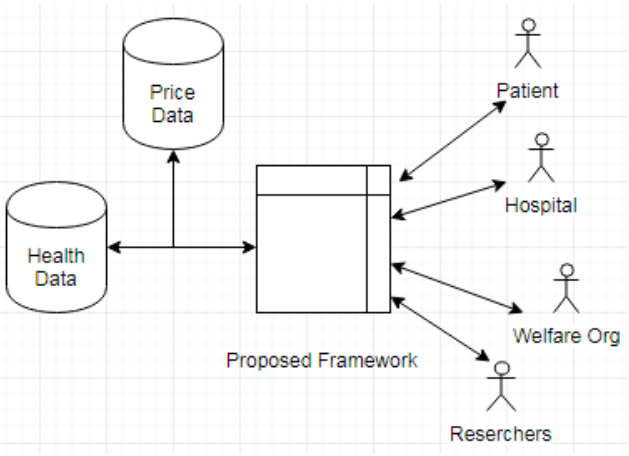


Figure-3. Proposed system architecture.

Health data: This data refers to full-fledged data related to health services provided and charges of particular services. Data size may varies from megabytes to gigabytes.

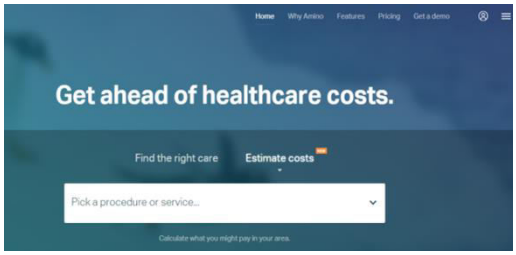
Price data: This data is the cost of each service provided by healthcare organization. This data is real time data and collected from hospitals.

Stakeholders: Patient: Patient is key stakeholder who can get all the required information from the system such availability of service, quality of service, and cost of the service.

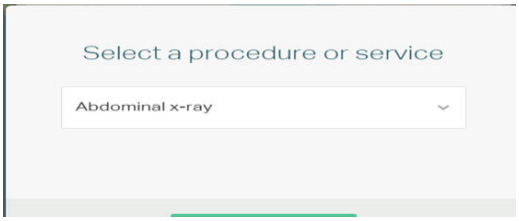
Hospital: Hospital those who wants exposure and wants there facility to be utilized by patient can be benefited by this system.

Welfare organization: Welfare organization can get benefit from this system in the form of availability of the system and concession available for the system.
Performance Analysis:

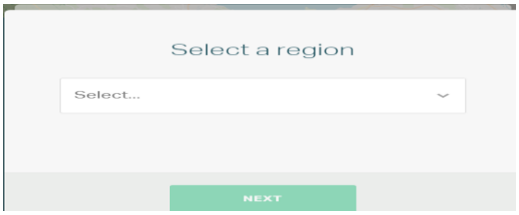
SAMPLE DESIGNS



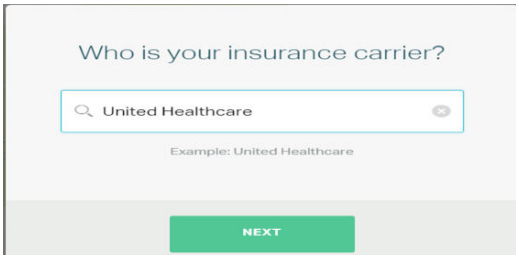
Design 1: Selection of service.



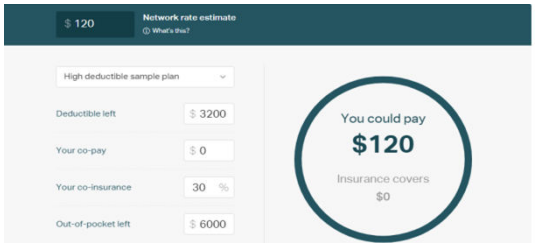
Design 2: Selection of procedure.



Design 3: Selection of region.



Design 4: Selection of insurance carrier.



Design 5: Summary of results.

Classification matrix is a vital tool for assessing the results of prediction as a result of it makes it straightforward to grasp and account for the consequences of wrong predictions. By viewing the number and percentages in every cell of this matrix, it shows however typically the model expected accurately.

Classification matrix	**0** (Actual)	**1** (Actual)
0 (Predicted)	True Negative	False Positive
1 (Predicted)	False Negative	True Positive

True negative: The person doesn't have the guts illness and therefore the take a look at is negative means that it's classified as negative.

False positive: The patient doesn't have the guts illness however the take a look at is positive means that it's classified as positive.

False negative: The patient has the guts illness however the take a look at is negative means that it's classified as negative.

True positive: The patient has the guts illness and therefore the take a look at is positive means that it's classified as positive.

Sensitivity: Measures the extent of genuine positives that region unit appropriately known all by itself

$$Sensitivity = \frac{TP}{TP + FN}$$

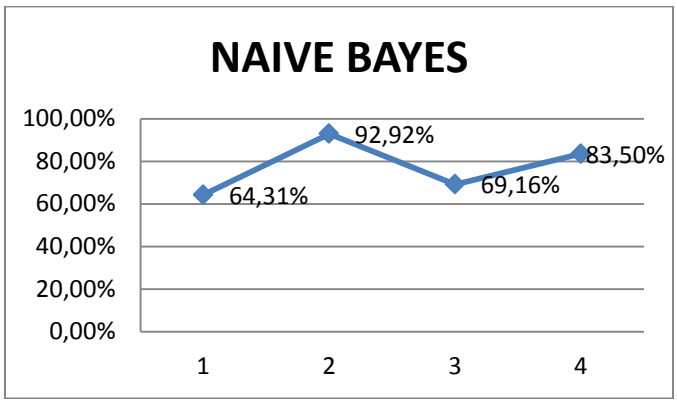
Specificity: Measures the extent of negatives that are properly known in and of itself

$$Specificity = \frac{TN}{TN + FP}$$

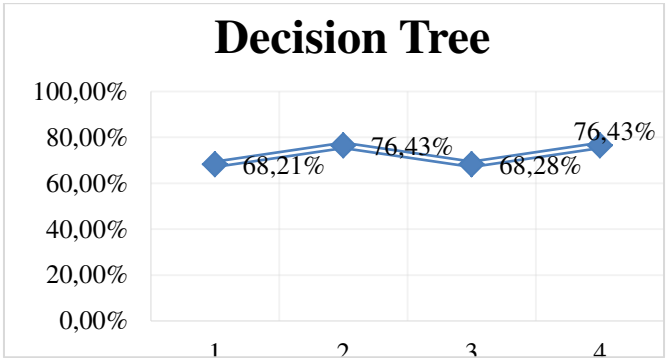
Accuracy: the accuracy is that the extent of genuine outcomes (both genuine positives and genuine negatives) inside the population.

$$Accuracy = \frac{TP + TN}{TP + FN + TN}$$

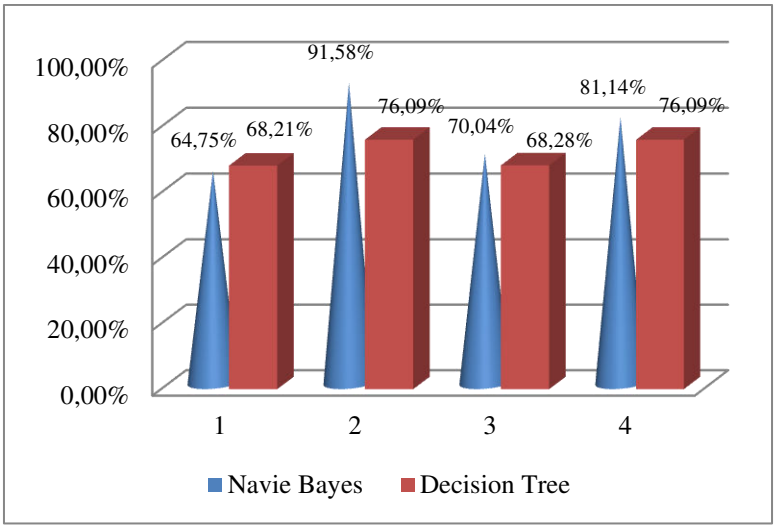
Sample experiment is performed on UCI & Actual data set collected from hospital.



Graph-2. Performance of Naïve Bayes.



Graph-3. Performance of decision tree.



Graph-4. Combined graph for Naive Bayes & decision tree.

6. CONCLUSIONS

From above proposed system healthcare system can be benefited if we apply machine learning techniques for predication. We have identified following research objectives in general research objectives defines what we can expect to achieve by a project.. In general research objectives defines what we can expect to achieve by a project .Future work can be study the criteria used by the patients in choosing the hospital service providers & to examine the level of satisfaction of patients and the factors influencing their level of satisfaction and to evaluate the quality of health care services provided by the private hospitals & to offer suggestions in enhancing the quality of hospital services and patient satisfaction.

REFERENCES

[1] Miguel, 2018, Using Machine Learning to Improve the Predication of Functional Outcome in Ischemic Stroke Patients.

[2] Juliana, Ivancakova, Frantisek Babic, 2018.Comparison of Different Machine Learning Methods on Wisconsin Dataset. IEEE World Symposium on Applied Machine Intelligent and Informatics.

[3] Venketesh Palanisamy, Ram Kumar Thirunavukarasu.,2017,Implications of big data analytics in developing healthcare framework- A Review, Journal of King Saud University Computer and Information Sciences.

[4] PVRD Prasad Rao and E. Rudrani. 2014. Privacy Preserving for Hospital Data Using Earth Mover's Distance International Journal of Applied Engineering Research in May 2014 No 15, pp. 2799-2808.

[5] Ruogu Fa, Samira Pouyanfar. 2016. Computational Health Informatics in the Big Data Age: A Survey in ACM.

[6] Junfeiqiu, Qihui Wu, Guoru Ding, Yuhuaxu. 2016. A Survey of Machine Learning for Big Data Processing. In Springer.

[7] Jianguo chen, Kenlili, Zhuotang, Kashifbilal, Keqin Li. 2016. A Parallel Patient Treatment Time Algorithm and Its Application in Hospitals Queuing Recommendation in Big Data Environment. In IEEE.

[8] Javier Andreuperez, Carmen C.Y Poon, Robert. 2015. Big Data for Health. In IEEE.

[9] Asif Adil, Hushmat Amin, Rajendra Jangir, Shabir Ahmed. 2015. Analysis of Multi Disease Using Big Data for Improvement in Healthcare. In IEEE Conference.

[10] Nina S. Godbole, John Lamb. 2015. Using Data Science & Big Data Analytics to Make Healthcare Green. In IEEE.

[11] Reena Duggal, Sunil Kumar, Balvindershukla. 2015. Improving Patient Matching: Single Patient View for Clinical Decision Support Using Big Data Analytics. In IEEE.

[12] Marco Viceconti. 2014. Peter Hunter Rose Hose. Big Data, Big Knowledge: Big Data for Personalized Healthcare in IEEE Journal.

[13] Xue Wen Chen, Xiao tong lin. 2014. Big Data Deep Learning: Challenges and Perspectives. In EEE.

[14] Maryam M Najafabadi, Flaviovillanustre, Taghi M Khoshgoftaar, Naeemseliya, Randall Wald1and Edinmuh aremagic. 2015. Deep Learning Applications and Challenges in Big Data Analytics. in Springer.

[15] Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao and Athanasios V. Vasilakos. 2015. Big Data Analytics: A Survey. in Springer.

[16] Sara Landset, Taghi M. Khoshgoftaar, Aaron N. Richter and Tawfiqhasanin. 2015. A Survey of Open Source Tools for Machine Learning with Big Data in the Hadoop Ecosystem. in Springer.

[17] Jing Lu, Malcolm keech. 2015. Emerging Technologies for Health Data Analytics Research. In IEEE.

[18] Hibaasri, Hajarmousaanif, Hassanal, Thomas Noel. 2015. Big Data in Healthcare; Challenges and Opportunities. In IEEE.

[19] Ashwin Belle, Raghu ram Thiagarajan. 2015. Big Data Analytics in Healthcare. In Hindawi Publishing Corporation.

[20] Seth Earley. 2014. Big Data and Predictive Analytics: What's New? In IEEE.

[21] Harsh Kupwade Patil, Raviseshadri. 2014. Big Data Security & Privacy Issues in Healthcare. In IEEE Computer Societ.