ISSN 2395-1621

A Review on Medical Big Data Analysis and Classification using the Data Mining Technique

[#1]Supriya Biradar, [#2]Prof. S. T. Waghmode

biradarsupriya207@gmail.com
stwaghmode@gmail.com

[#12]Department of Computer Engineering
Imperial College of Engineering and Research, Wagholi, Pune.

ABSTRACT

Now day's big data processing is most important for any platform and critical also so, we analysis using the data mining technique is a new methodology that is introduced in the field of computer science to abstract the size of data. The proposed system uses to large quantities of data which need special processes according to business analysis. The user to get correct proper information and gain accurate knowledge for the future prediction also this all gain using the data mining technique. Now days a lot of people lost their life because of medical mistakes. This proposed system discusses big data mining technique in medical dataset and we also analyze this clinical big datasets to discover knowledge to use it in clinical prediction to the hospital. Afterward, this system propose framework which figure-out how to analyze big data using data mining and how to discover knowledge from the extracted information. We also classification like area-wise, age-wise, dieses-wise according to patient criteria check medical prediction for future care about that medical treatment. As conclusion of the big data analysis is expected to the knowledge structure which guided the decisions making for in future predication.

Keyword: Data Mining, Big data analysis, Health record, patient record, Classification.

ARTICLE INFO

Article History

Received: 25th May 2018

Received in revised form :

25th May 2018

Accepted: 28th May 2018

Published online :

28th May 2018

I. INTRODUCTION

Data Mining is a technique to analysis unstructured data to structured data process which designed to explore data in search of given patterns and/or relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The main goal of data mining is prediction of given dataset and predictive analysis. Data mining is the most common type of data mining and one that has the mostly uses business applications. The process of data mining has three stages:

(1) Data exploration,
(2) Model building or pattern identification with validation/verification, and
(3) Deployment

Data mining Applications are mostly used where the users can send the feedback to the any application, Data collection has increased widely and is beyond the capability of commonly used software tools to capture, manage, and process within a "tolerable elapsed time." The most challengeable part for Big Data applications is to explore the large volumes of data and extract useful information or knowledge for future actions. The big data extraction process has to be very efficient as well as easy and close to real time because storing all observed data is nearly infeasible.

Data is being produced increasing rate so critical to analysis. There has also been acceleration in the proportion of machine-generated and unstructured data compared to structured data such that 80%. All data holdings are now unstructured and new approaches upcoming so mining technique is best option to convert structure data for users identification.

PROBLEM DEFINITION:

Today increasing number of organizations is facing the problem of explosion of data and the size of the databases used in today's enterprises has been growing at exponential rates. Big data is upcoming through many sources application like business processes, online transactions,

social networking sites like Facebook, twitter, web servers, etc. and remains in structured as well as unstructured form. Today's business applications are having enterprise features like large scale, data-intensive, web-oriented and accessed from diverse devices including mobile devices.

Processing or analyzing the huge amount of data or extracting meaningful information is a challenging task. The term "Big data" is used for large data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes on servers are a constantly increases few dozen terabytes to many peta bytes of data in a single data set.

II. REVIEW OF LITERATURE

Many researches have been done on the prediction of data mining technique using the different technique and different scenario for data mining analysis.

[1] Ying Dai, Jin Tian, Hao Rong, Tingdi Zhao, "Hybrid safety and analysis method based on SVM and RST", in this research paper focused on providing a safe landing without an accident using the support vector machine on aircraft dataset.

[2] A.B Arockia Christopher and Dr. S.Appavu "Data mining approaches for aircraft accidents prediction", this paper used decision tree method to predict and analysis the warning level. Dataset used are pilot details, delay details, accident related details, maintenance details and flight details.

[3] A.B Arockia Christopher and Dr. S.Appavu in this paper analyzed various data preprocessing techniques to find best techniques which suits for airline dataset. In data mining classification algorithms and clustering techniques are used in for comparison. In that paper the data mining tool weka was used in this process. The results of this analysis they have proved that better mining technique transformer would perform better than other attribute evaluators on airline data to reduce the dataset. The study of data mining technique on Turkey airline, here also decision tree technique is used to generate model. This model in turn is used to predict the warning level.

[4] ZohrehNazeri, George Donohue, Lance Sherry, "A similar research, "Analyzing Relationships between Aircraft Accidents and Incidents", this paper study was done in USA. In this research paper various accidents and causes are analyzed for happening these accidents. All accident details from NTSB database and reasons for these accidents are maintained. The taxonomy was used for filtering the upcoming data. In order to maintain an unstructured data to structure data transformation is applied to transform the report into a vector containing common fields. Then author used STUCCO algorithm as used for finding the pattern for future analysis. The result from the finding is then ranked using factor support ratio measure. Accuracy level of the output produced is also determined using accuracy algorithm.

[5] D.K.Y Wong, D.E. Pitfield, R.E Caves and A.J Appleyard, "The development of aircraft accident frequency model", used weather as the only parameter. Temperature level, humidity, storm and wind speed are used as data set. Logistic regression analysis is used to estimate accident probability in a given weather condition.

In a research paper "A system approach to accident causation in mining", [6] Michael G. Lenne, Paul M. Salmon, Charles C. Liu, Margaret Trotter, analyzed human factors and classification systems (HFACS) were used to rise and caution level to any kind of accidents. Dataset from various accidents are stored in the database. Human error, technical faults, natural environment and climate conditions datasets were used.

III. SYSTEM OVERVIEW

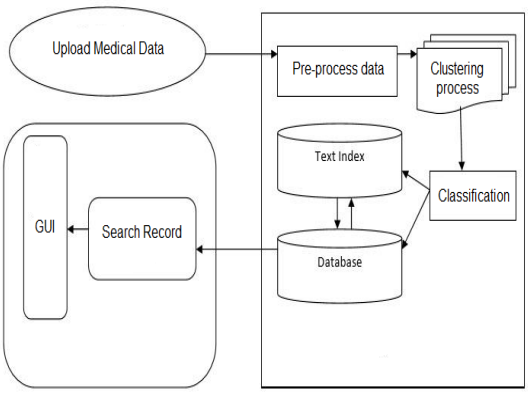

Fig 1. System architecture

Above fig1. Shows the system architecture which include all process as well as models to performing using data mining technique.

A. Admin:

Admin is responsible for overall analysis the given input dataset. The fired the queries according to his need. After that the server System is responsible for the processing the mining further. Finally admin get the analysis result according to system scenario.

B. Clustering:

Here data is clustered for grouping upcoming data to easily analysis. Below steps shows the different stages during the clustering.

- Medical data Partition into given k non-empty set.
- Identification of cluster for grouping.
- Assign each point to a specific cluster.
- Result generate grouping clustered.
- Repeat the above steps for re-allotted points and find the mean point for the new cluster.

C. Naïve Bayes Algorithm:

Naïve Bayes Algorithm is used for the classification as well as pattern analysis. Used algorithm gives better output, based on Bayes theorem and frequency table. It gives the

Estimation, Classification, and Prediction. It is used when large data set. It is very easy to construct. Not using complicated iterative parameter estimations.

IV. CONCLUSION

We conclude big data analysis on medical dataset. The presence of big data has produced a unique moment in the history of data analysis. In this paper, we provide a detailed comprehensive study of the data mining techniques, analyzing the new approach and new scenario that have been introduced to some of them that have been successfully developed into big data analytic techniques.

V. ACKNOWLEDGMENT

I wish to express my profound thanks to all who helped us directly or indirectly in making this paper. Finally I wish to thank to all our friends and well-wishers who supported us in completing this paper successfully I am especially grateful to our guide Prof. S. T. Waghmode Sir for him time to time, very much needed valuable guidance. Without the full support and cheerful encouragement of my guide, the paper would not have been completed on time.

REFERENCES

[1] Jin Tian, HaoRong, Tingd Zhao, "Hybrid Safety analysis method based on SVM and RST: An application to carrier landing of aircraft", School of Reliability and Systems Engineering, Vol. 80, Dec. 2015, Pages 56-65.

[2] A.B. Arockia Christopher, S. Appavu, "Data Mining Approaches for Aircraft Accidents Prediction", Emerging Trends in Computing, Communication and Nanotechnology, 2013, Pages 25-26.

[3] A.B. Arockia Christopher, S. Appavu, "Feature Selection for Prediction of Warning Level in Aircraft Accidents", Advanced Computing and Communication Systems (ICACCS), 2013, Pages 1- 6.

[4] ZohrehNazeri, George Donohue, Lance Sherry, "Analyzing Relationships between Aircraft Accidents and Incidents", International Conference on Research in Air Transportation (ICRAT), 2008 Pages 185-190.

[5] D.K.Y Wong, D.E Pitfield, R.E Caves and A.J Appleyard, "The Development of Aircraft Accident Frequency Models", Safety and Reliability for Managing Risk – GuedesSoares, 2006 Pages 83-90.

[6] Michael G.Lenne, Paul M. Salmon, Charles C. Liu, Margaret Trotter, "A System Approach to Accident Causation in Mining", Accident Analysis and Prevention. Vol. 48, Sep 2012, Pages 111-117.

[7]http://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx - For dataset reference.

