

An Exploration of Various Data Mining Techniques for Application in Child Healthcare

Abhijeet Sudhakar, Rajendra B. Patil, Srivaramangai R



Abstract: Data mining has been predominantly used for analysis of various data and findings. It has already gained ample importance in the field of medicine especially the healthcare department of governments. It is a useful tool for analyzing the child health care data for government so that enough measures can be taken to reduce the causes of deaths. This paper aims at exploring different data mining techniques that have been implemented especially in the child healthcare sector in order to get analytical data for decision making. Various techniques have been used for finding the mortality rates among infants, the under nutrition percentage, the causes of different diseases among children of under 5 years age. This paper gives an analysis of the earlier research and presents the need for developing new algorithms for healthcare with more accuracy in predictions.

Keywords : Data mining, Healthcare, Data Analytics, Artificial Neural Networks, Naïve Bayes Classifier, Support Vector Machine, K-Nearest Neighbour

I. INTRODUCTION

Healthcare sector is a major area where we have lots of critical data available which need analysis and inferences. For any region, state, country, health is an essential part of life which cannot be ignored. In world statistics there is a separate place for healthcare. This industry has lots of big data such as information on diseases, age, gender, nutrition, child health etc. We focus on child healthcare. It focuses on child birth, still born, diseases, vaccination, malnutrition and so on. So we have lots of data for this. In order to improve the child healthcare, there is a need for evaluating these data for decision making by the officials of healthcare sector and government. But studying this large amount of data is tedious using conventional means. Due to diverse scope of volume of data it becomes difficult to use the traditional analytical methods. Data mining techniques are widely used in making decisions by analyzing the big data. The intelligent techniques of data mining helps in extracting the hidden data and also brings to limelight the unidentified data which may be of great significance in decision support systems.

Manuscript published on 30 September 2019

* Correspondence Author

Abhijeet Sudhakar, Student, University of Mumbai, Mumbai, India.
Email: abhijeet.sudhakar@gmail.com

Rajendra B. Patil, Assistant Professor, S.K.Somaiya College of Arts, Science & Commerce, Mumbai, India. Email: patilrajendrab@gmail.com

Srivaramangai R, Assistant Professor, Department of Information Technology, University of Mumbai, Mumbai, India. Email: rsrimangai@udit.mu.ac.in

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

II. DATA MINING CONCEPTS & TECHNIQUES

Data mining is a technique which is generally applied to extract information from voluminous data, analyze that and do a pattern matching with the historical data in order to

predict or forecast the possibility of occurrence of future events. The analysis is done in three ways- Descriptive analysis, Predictive analysis and Prescriptive analysis. Figure 1 describes these three categories along with one added category which is often mentioned in most of the research articles as diagnostic analysis.

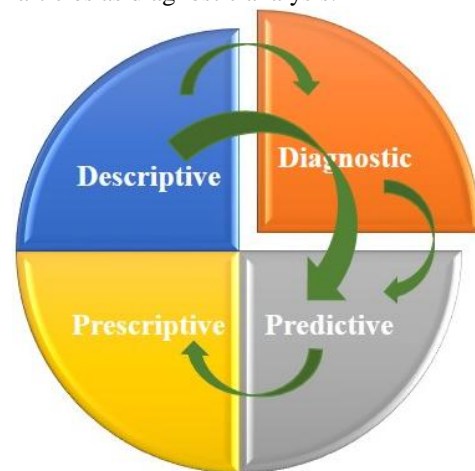


Figure 1. Categories of Data Analysis

A. Descriptive Analysis: It's a term given to the assessment of data that depicts, show up or condense data in a significant way to such a degree, that patterns may ascend out of the data.

B. Predictive Analysis: A techniques which is used to predict the occurrence of unknown events.

C. Prescriptive Analysis: It is a techniques used to find best strategy for a given circumstance

Diagnostic analysis is sometimes included in the categories of analysis where it gives the reasons for certain happenings after examining the given data. Prescriptive analysis is an advanced level which includes the other three. In order to do the above types of analysis, various techniques have been developed and predominantly used for data analytics. Some of the most widely used techniques are regression, classification, prediction, outlier detection, association, clustering and pattern matching.

The algorithms which are mostly used for implementing the techniques are artificial neural networks, support vector machine, K means clustering, K Nearest Neighbour, Naive Bayes, Adaboost, Apriori etc. The survey done in this paper reveals the performance of these algorithms for various techniques focusing on healthcare. Further, it examines the implementation methods of these techniques by researchers and gives an overall performance of algorithms used in their research.

III. RATIONALE FOR CHILD HEALTH CARE

The infections and disorders caused to children under 5 years of age are a serious health issue in many countries leading to morbidity. The Data mining in healthcare is lagging behind compared to other areas as the size and diversity is humongous. The analysis of demographic and healthcare survey data and the inferences made out of it will become a support system for the healthcare department to take preventive and precautionary steps in reducing the mortality rate. For the past few years latest sophisticated techniques are used to implement data mining in healthcare. These techniques have enabled to the data to be transformed into valuable information and assist in decision making by deriving the knowledge hidden in the data. Once the necessary operations of analytics are done, the data is mined and the systems will be in place to make data mining discernments operational, thus paving way for use of predictive analytics in novel and inventive ways. The main purpose of selecting child health care for surveying the data mining techniques is to understand the best algorithms which results in presenting significant factors that affect kids. This study will help in understanding the level of research that has been already done and their performance in prediction. This knowledge can help in further investigations of methods to enhance the existing systems or to bring new ideas with accuracy levels which will enable data miners, policy makers, healthcare professionals to take the right decision in effective intervention and prevention.

IV. SURVEY AND ANALYSIS FROM EARLIER RESEARCH

Ramageri[1] talks about the data mining techniques used in various business and its applications. It is a process to take large amounts of data to study. The various algorithms are-Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc. The various steps involved in this are- Exploration, pattern identification and deployment. Data mining has wide variety to applications. For business models, forecasting, pattern finding, knowledge discovery is important, it is achieved through data mining. Techniques like classification, clustering, etc are very useful for finding patterns which can be used to predict the future for businesses. Thus data mining has got its uses in various fields and is up and coming concept in field of databases and information. Jothi et al[2] have observed that KDD has lots of inputs about handling data and one of them is data mining. Data mining in healthcare is a huge subject. There are many areas of studies and application using data mining in healthcare industry. There are different

disciplines like machine learning, AI, statistics and probability and models- predictive and descriptive. The data mining tasks include classification, association rules, clustering and anomaly description. They have reviewed various methods of data mining are SVM, decision tree, k nearest neighbor etc. Healthcare sector extensively use data mining techniques. They use it for prediction, diagnosis of different diseases and new methods for treatment of these diseases. Since many techniques are used, their accuracy rates are different. In order to get the best results i.e. accurate results, hybrid models are implemented. Arun and Ali[3] Data mining is algorithmic technique for getting new useful patterns of data. Here it focuses on healthcare industry. Data mining has become popular in this sector. We use KDD and data mining. We have supervised and unsupervised learning. The techniques used are decision tree, neural networks and naïve bayes classifier. We also have clustering, associative rule, k nearest neighbor etc. Thus information on new ways of saving life is obtained. The medical experts use this on their previous data base and get more useful and latest information. Medical authorities can use data mining for their medical policies like vaccination, sterilization etc. They use data mining to predict spread of diseases or advent of new disease and take preventive measures. Business Intelligence is used for diagnosis for diseases. This paper says that data mining techniques are very widely used in medical data for diagnosis and investigation of several diseases. Oswal and Shah[4] This paper tells us about the various studies of data mining in health issues and its applications in health sector. We have descriptive analysis, prescriptive analysis and predictive analysis in data mining evaluation. The techniques used are decision tree, Naïve Bayes and neural networks. It has various uses in CAD, ALL(Acute lymphatic leukemia), various biomedical applications, in hospitals and so on. Healthcare industry relies on data mining for disease prediction and diagnosis. This paper gives information on current data mining practices used in healthcare sector. Bekri and Govardhan[5] The demand and want for data mining is more in the area of healthcare, regardless of variations and conflicts in processes. Various discussions led to the demand of data mining in the field of healthcare which includes both public health as well private health. From previous records in computer, sufficient facts are obtained. The data is humongous in size, so one person has difficulty in obtaining all the relevant information[5]. Specialists consider that the improvement in medicals has reduced leading to complication of recent medical data. To overcome this drawback computers and data mining can be utilized. The research of data mining utilization in medicine and public health offered only synopsis of present observations and disputes. Healthcare institutions and firms would utilize this method of data mining to get more details and knowledge from the information that is already present in their institution records. An institution must mention all rules and strategies on the safety and confidentiality of victim's data, before boarding on data mining. The same rule must be said to its partners and implied to other institutions and branches.

Public health related issues like pandemic occurrence, the want to identify the commencement of disease in a non-invasive, easy manner and the want to be more reactive towards the patients. All these factors must be considered important and the desire for health institutions to combine information and data mining must be utilized to examine this information. Sheenal and Hardik[6] In this paper we take step by step approach. First is selection, then preprocessing, transformation, data mining and result or evaluation. The techniques are classification, clustering and association. We use data mining for treatment of various diseases.

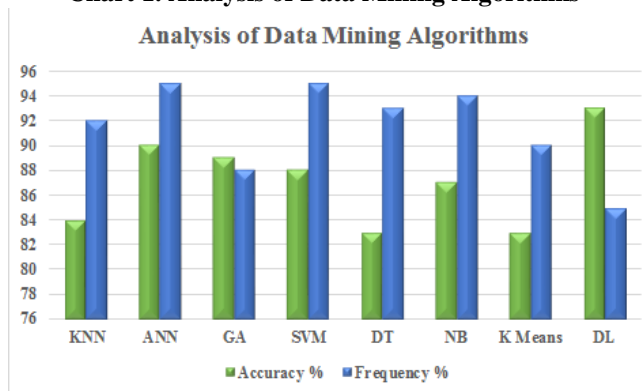
Gharehchopogh et al [7] talk about data mining as a process in which new patterns, data sifting, patterns, trends, correlations, etc are done using statistical and mathematical techniques (Gartner Group). The techniques are kd method and decision tree algorithm. Artificial intelligence and statistical tools use large amount of data which is discovered by data mining. It has become very successful in bpth scientific as well as business communities. People’s behavior, medical data, client services etc are all done by data mining. Even Alzheimer and Estrogen relationship was discovered using data mining. Sanati-Mehrziy et al[8] Artificial Intelligence, Databases, Statistics have all one thing in common i.e. data mining. Though it is relatively new concept, it has seen an increase in demand all over the world. Especially Healthcare sector benefits significantly from data mining. It can discover about people living in some part f world, their diet, their lifestyle and diseases and comparison with other group of people. Many business organizations use data mining. The various applications are hearing aid dataset, cancer cell mining, DNA speculation etc. Durairaj and Ranjani[9] This paper gives overall view of data mining. It defines data mining. Then we see the applications in various health sectors and the results are produced which are recorded. Different disease analysis is done. Peter et al.[10] have evaluated the data sets they obtained from northern Thailand and found that Akaike information criterion (AIC) and Bayesian information criterion (BIC) techniques gave a better predictability. Niaksu et al.[11] have done a review of the data mining techniques that have been used in medical field and drawn conclusions with the data and references given by PubMed annotated MeSH terms. Tanvi et al.[12] have analyzed the effectiveness of data mining techniques in health care data and they have compared the performance of various approaches, tools in order to recommend an efficient method that can be used in this sector. M. H. Tekieh and B. Raahemi[13] in their analysis have found that evaluating the health data from various sources like electronic health record(EHR) has led to the necessity of using data mining techniques due to its complexity and specific challenges. Herland et al.[14] have done a review of health informatics using BigData and suggested that narrowing to one or more techniques as best is difficult since this area needs a lot of testing and confirmation on the outputs before it could be applied on the prospective data. The use of BigData is indeed an advantage to health informatics as it can deal with large amount of variety, velocity, veracity etc. Priyanka et al.[15] have presented a new methodology in data mining for predicting the probable diseases based on user input symptoms. It also provides remedial solutions as an assistive system to doctors and

suggestions for the users to live a healthy life. Bekri and Goavardhan[16] have done a research on the use of data mining techniques for healthcare and have put forward the important factors to be considered while applying these techniques. Soni and Gandhi[17] have done a survey of usage of data mining techniques in medicine and healthcare and the challenges faced by these techniques. They suggest that the domain knowledge from experts is very important for the success of these techniques and so the medical professional need to be taken into confidence while implementing such techniques. Yadav et al.[18] in their survey have found the applications of data mining algorithms like decision tree, artificial neural networks, genetic algorithms and k nearest neighbor. They have analyzed the use of these techniques for various medical problems. Each algorithm has got unique application in many of the detections of diseases and disorders. Fallah et al.[19] have done a review of research work done on healthcare using data mining and have listed out the algorithms that has been used in each of the research work. It gives year wise papers published and the accuracy of the algorithms in detecting the diseases. Monitoring, predicting, detecting, classifying, data gathering and recommendations are the functions which has been dealt by Apps developed for mobiles using data mining. Providing an organized way of enhancing the ability of health domain experts thus availing the support and confidence of the patients to handle their own health issues in an effective way is the primary aim of healthcare organizations for which efficient data mining techniques have been stated by Panagioti et al.[20]. Due to the efficient way of classification support vector machines have been used more than other techniques. Obviously it gave better accuracy levels in prediction as per the work done by Pouladzadeh et al.[21].

V. OBSERVATIONS AND CONCLUSION

This paper gives an overview to data mining and the different techniques thus presenting a report on the survey done on these techniques. After understanding the basic concepts of data mining, it reviews both descriptive and predictive analysis done using predominant techniques from the previous work done by researchers on medical domain especially healthcare.

Chart 1. Analysis of Data Mining Algorithms



The chart 1 gives the analysis of accuracy and frequency of usage of prominent data mining techniques like K nearest neighbor(KNN), artificial neural networks(ANN), genetic algorithms(GA), support vector machines(SVM), decision tree(DT), Naïve Bayes(NB), K means clustering (K Means), deep learning(DL) on healthcare sector. This survey enables us to understand the demand for data mining techniques and the scope of research before us to enhance the accuracy level of these above mentioned technique. It also stimulates to develop new algorithms which may give more accuracy and correct predictions for this area, as medical field needs very cautionary methods to be used before implementing on live data.

The authors could learn the application of data mining techniques in healthcare sector and also the usage of data mining algorithms. This exploration also gave an understanding that more testing is required with retrospective data of larger volumes which paves way to continue to the next level of application of algorithms in child healthcare data which has been downloaded from the Government of India healthcare unit website.

REFERENCES

1. Bharati M. Ramageri, "Data Mining Techniques and Applications", Indian Journal of Computer Science and Engineering, Vol. 1, No. 4, 2011, pp. 301-305.
2. Neesha Jothia , Nur'Aini Abdul Rashidb , Wahidah Husainc , "Data Mining in Healthcare – A Review", Proceedia Computer Science, Elsevier Open Access Article Under The CC BY-NC-ND License 72, 2015, pp.306-313.
3. Arun Pushpanl, Ali Akbar, "Data Mining Applications in Healthcare", IOSR Journal of Computer Engineering (IOSR-JCE), 2017, pp.04-07.
4. Sangeeta Oswal , Gokul Shah, "A Study on Data Mining Techniques on Healthcare Issues and its Uses and Application on Health Sector", International Journal Of Engineering Science and Computing, Vol.7, No.6, 2017, pp.13536-13538.
5. Fawzi Elias Bekri, Dr. A. Govardhan, "Association of Data Mining and Healthcare Domain: Issues and Current State of the Art", Global Journal Of Computer Science and Technology, Vol. 11, No. 21, 2011, pp. 1-8.
6. Sheenal Patel, Hardik Patel, "Survey of Data Mining Techniques Used in Healthcare Domain", International Journal of Information Sciences and Techniques (IJIST), Vol.6, No.1/2, 2016, pp. 53-60.
7. Farhad Soleimanian Gharehchopogh, Peyman Mohammadi, Parvin Hakimi, "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study", International Journal of Computer Applications, Vol. 52, No.6, 2012, pp. 21-26.
8. Dr. Reza Sanati-Mehrizy, Jeffrey H Wrigh, Dr. Afsaneh Minaie, Dr. Ali Sanati-Mehrizy, Paymon Sanati-Mehrizy, "A Study of Application of Data Mining Algorithms in Healthcare Industry", American Society for Engineering Education, 2013
9. M. Durairaj, V. Ranjani, "Data Mining Applications in Healthcare Sector: A Study", International Journal of Scientific & Technology Research, Vol. 2, No. 10, 2013, pp.29-35.
10. Peter Haddawy Myat Su Yin, Tanawan Wisanrakkit, Rootrada Limsupavanich, Promporn Promrat, Saranath Lawpoolsri, Patiwat Sa-Angchai, "Complexity-Based Spatial Hierarchical Clustering For Malaria Prediction", Journal Of Healthcare Informatics Research, Vol.2, No.4, 2018, pp. 423-447.
11. Niaksu O., Skinulyte J., Duhaze H.G. (2014) A Systematic Literature Review of Data Mining Applications in Healthcare. In: Huang Z., Liu C., He J., Huang G. (eds) Web Information Systems Engineering – WISE 2013

- Workshops. WISE 2013. Lecture Notes in Computer Science, Vol 8182. Springer, Berlin, Heidelberg.
12. Tanvi Anand, Rekha Pal, Sanjay Kumar Dubey, "Data mining in healthcare informatics: Techniques and applications", 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2016, pp. 4023-4029.
13. M. H. Tekieh, B. Raahemi, "Importance of data mining in healthcare: A survey," 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Paris, 2015, pp. 1057-1062.
14. Matthew Herland, Taghi M Khoshgoftaar , Randall Wald , "A review of data mining using big data in health informatics", Journal Of Big Data, Vol.1, No. 2, 2014.
15. Priyanka, Sana Khan, Tulsi Kour, "Investigation on Smart Health Care Using Data Mining Methods", Section: Research Paper, Product Type: Isroset-Journal, Vol.4 , No.2 , 2016, pp.31-36.
16. Fawzi Elias Bekri, Dr. A. Govardhan, "Association of Data Mining and Healthcare Domain: Issues and Current State of the Art", Global Journal of Computer Science and Technology, Vol.11, No. 21, 2011, pp. 1-8.
18. Nakul Soni, Chirag Gandhi, "Application of Data Mining to Health Care", International Journal of Computer Science and its Applications, 2005, pp.245-248.
19. Dr. Surendra Kumar Yadav, Nitesh Dugar, Aditi Jain, "Future Perception In Public Health Care Using Data Mining", International Journal of Engineering Research & Technology (IJERT), Vol. 3, No. 1, 2014, pp. 1246-1250.
20. Mina Fallah, Sharareh R. Niakan Kalhori, "Systematic Review of Data Mining Applications in Patient-Centered Mobile-Based Information Systems", Health Informatics Research, Vol.23, No.4, 2017, pp.262-270.
22. Panagioti M, Richardson G, Small N, Murray E, Rogers A, Kennedy A, "Self-management support Interventions to reduce health care utilisation without compromising outcomes: a systematic review and metaanalysis" BMC Health Serv Res 2014;14:356
24. Pouladzadeh P, Kuhad P, Peddi SV, Yassine A, Shirmohammadi S., "Mobile cloud based food calorie measurement", Proceedings of 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2014, pp.1-6

AUTHORS PROFILE



Abhijeet Sudhakar is a student of Thadomal Shahani Engineering College, University of Mumbai, Mumbai, India who has graduated his Bachelor of Engineering in Electronics and Telecommunications and has an interest in the area of data science. He is working under the guidance of Dr. Rajendra Patil and Dr. Srivaramangai for research projects on data mining for healthcare.



Rajendra B. Patil is heading the department of Information Technology in S.K.Somaiya College of Arts, Science and Commerce, University of Mumbai, Mumbai, India and a PhD in Computer Science. He has several publications to his credit and his area of specialization is Data Mining, Data Science and Artificial Intelligence.



Srivaramangai R is heading the department of Information Technology at the University Department, University of Mumbai, Mumbai, India and a PhD in Computer Science. Her area of specializations are digital image processing, data analytics, Security and has several publications to her credit.