iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** IV **Month of publication:** April 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Predictive Analysis of Diseases: An Overview

Purvashi Mahajan[#1], Abhishek Sharma[#2]
M.Tech[#1], Assistant Professor[#2]
Department of Computer Science, Shri-Balwant Institute of Science and Technology, Dcrust University, Murthal

Abstract— Predictive Analysis of Diseases is an important issue in medical research. Majority number of patients struggle for their check up even for predictive diseases like heart attack, liver damage, eczema and diabetes. There is a lot of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. This research paper put lights on how predictive analysis of diseases play a vital role in healthcare domain and analyse various predictive data mining techniques. With the effective analysis tools and algorithms, pattern of diseases can be possible.
Keywords—Predictive Analysis, Healthcare, Data Mining, Heart Disease, Medical Diagnosis

I. INTRODUCTION

Predictive analysis of diseases has been a subject of research. Predictive Analysis plays a vital role in Healthcare industry. Medical data mining has great potential for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized for clinical diagnosis. Since the available data is heterogeneous in form and very huge, data needs to be organized. Data mining technology gives a user-oriented approach to hidden patterns in data.

According to an estimation of World Health Organization, 12 million deaths occur worldwide every year due to the Heart diseases. Half the deaths in the United States and other developed countries occur due to cardio vascular diseases. Numerous developing countries have fallen prey to cardio vascular disease.

Predictive analysis of diseases is considered as an important yet complicated task that needs to be executed accurately and efficiently. The automation of this field would be highly helpful. Since all the doctors do not possess forte in every subject and moreover there is scarcity of resource persons at certain places. Hence, an automatic medical diagnosis system would probably be helpful by bringing all of them together.

Proper and efficient implementation of automated system needs a comparative study of various techniques and algorithms available. This research paper aims to analyse the various predictive data mining techniques proposed in recent years for the diagnosis of diseases especially heart disease.

II. LITERAURE REVIEW

This section provides a review of Predictive Analysis of diseases using data mining techniques. An overview of usefulness of predictive analysis of diseases in healthcare domain is summarized below:

A. S. Vijayarani et al [1] suggested An efficient clustering algorithm for predicting diseases from hemogram blood test samples. The paper primarily focussed on predicting diseases from the hemogram blood test data set by using data mining techniques. Algorithms' performances are evaluated by using the cluster accuracy, error rate and execution time. From results, it is known that the proposed weight based k-means algorithm performance is better than other algorithms.

B. Jyoti Soni et al [2] presented Predictive data mining for medical diagnosis. The experiments performed to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and some time Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well.

C. B.Venkatalakshmi et al [3] focussed on Heart disease diagnosis using predictive data mining. A series of experiments were conducted with same data sets in Weka 3.6.0 tool. Data set of 294 records with 13 attributes is used and result revealed that Naïve Bayes outperforms and sometime Decision Tree. Genetic algorithm will be utilised in future to reduce the data size. As per comparison result, Accuracy of Naïve Bayes came out to be 85.03 whereas Decision Tree was 84.01.

D. R.Chitra et al [4] presented the Review of heart disease prediction system using data mining and hybrid intelligent techniques. The review paper concluded that data mining plays a major role in heart disease classification. Neural Network with offline training is a good for disease prediction in early stage and the good performance of the system can be obtained by pre-processed

©IJRASET: All Rights are Reserved

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

and normalized dataset. The classification accuracy can be improved by reduction in features.

E. Hlaudi Daniel Masethe et al [5] carried out a research on Prediction of heart disease using classification algorithms. Various classifiers were being during the research such as J48, REPTREE, Naïve Bayes, BAYES NET and SIMPLE CART. Predictive accuracy of J48, REPTREE & SIMPLE CART came out to be 99.0741, that of Naïve Bayes was 97.222 and BAYES NET was 98.1481. The predictive accuracy determined by J48, REPTREE and SIMPLE CART algorithms suggests that parameters used are reliable indicators to predict the presence of heart diseases.

F. Rebecca Hermon et al [6] presented Big Data in Healthcare: What is it used for. It explained the useful aspects of big data techniques in healthcare. Concluded was the impact of integration of data mining and medical informatics and its analysis using big data techniques on healthcare delivery costing and better heathcare results.

III. PREDICTIVE ANALYTICS PROCESS

A. Define Project

The outcomes of the project, deliverables, scoping of the effort and business objectives are defined. The data sets which are going to get used are being identified.

B. Data Collection

Data mining for predictive analytics prepares data from multiple sources for analysis. This gives a total view of the customer interactions.

C. Data Analysis

Data Analysis is a process of inspecting, organizing, transforming, cleaning and modelling data keeping in view the purpose of deriving useful and related information, making out the conclusions.

D. Statistical Analysis

Statistical analysis helps to verify or authenticate the assumptions, Pseudo statements, hypothesis and test them with the help of statistical models.

E. Predictive Modelling

Predictive modelling gives the ability to create accurate predictive models about future automatically. There are available options to choose the most effective solution along with multi model evaluation.

F. Predictive Model Deployment

Predictive model deployment gives the option to implement the analytical results in to the everyday decision making process for getting the results, reports and outcomes by automating the decisions based on modelling.

Various Data Mining techniques are used for Predictive analytics of Diseases. Such Techniques include K-means Cluster technique, A-priori algorithm, Particle Swarm Optimization, Decision Trees etc. These techniques are really helpful during stage 2 of Predictive analysis i.e. Data collection. Data collection stage requires data from various sources. For making that data organisable, data mining techniques play a vital role. Be it portioning data sets, organising data or predicting the future aspects of data, these techniques play a major role.

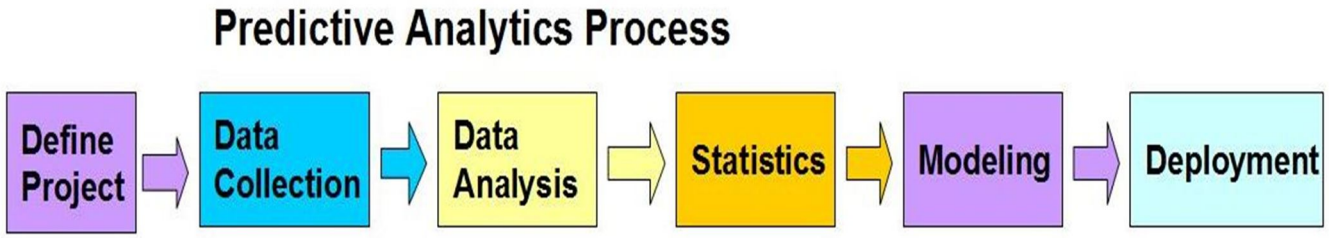


Fig. 1 Stages of Predictive Analytics Process

©IJRASET: All Rights are Reserved

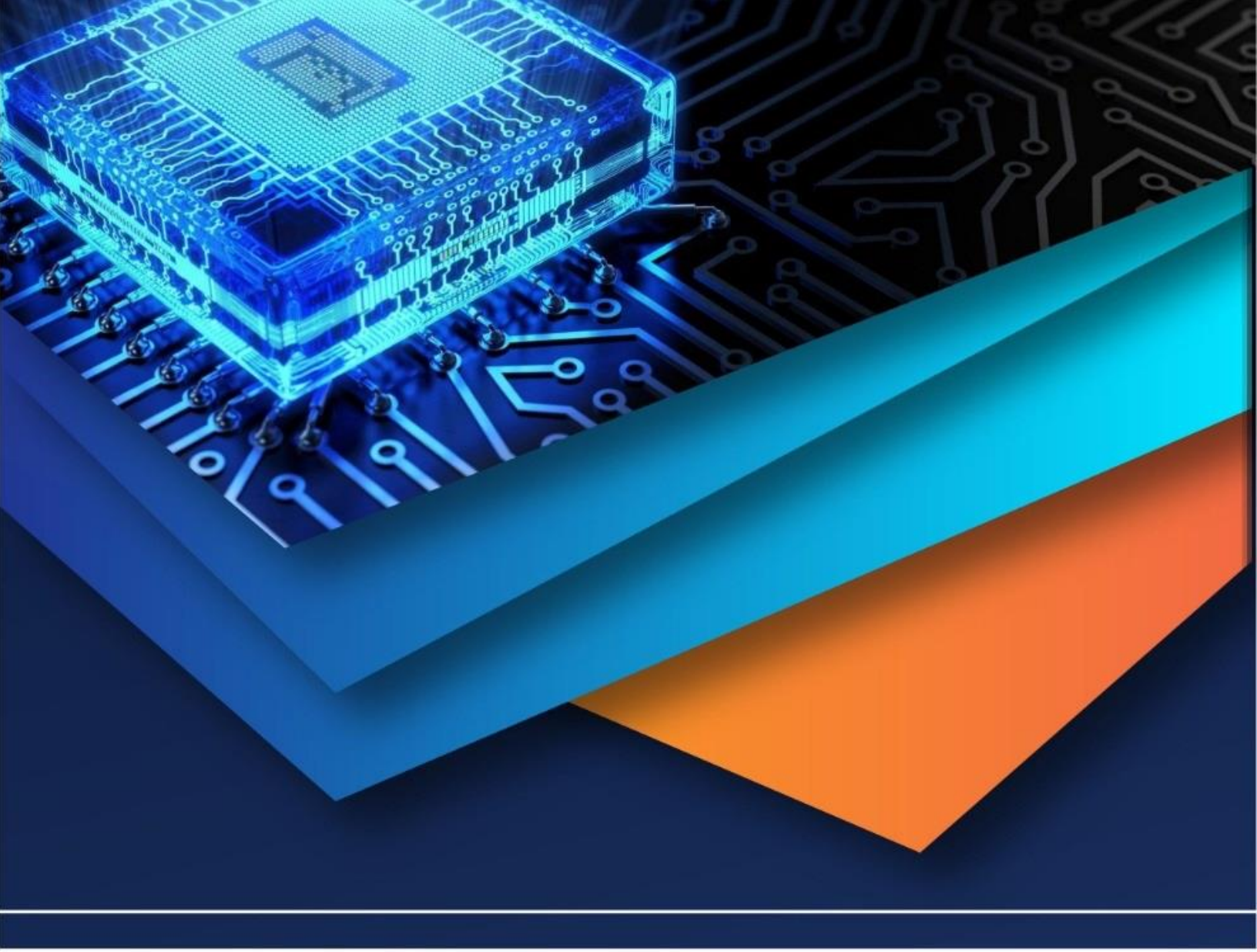
IV. CONCLUSION

The Data Mining techniques are quite helpful in one the stages of Predictive Analysis of Diseases i.e. Data Collection Stage where data needs to be prepared for predictive analysis before getting ready for next stage i.e. Data Analysis. These techniques provide organised, efficient and accurate data for further stages of the cycle. This is how data mining techniques play a vital role in Data Analysis Stage of Predictive Analysis.

REFERENCES

[1] R.Chitra and V. Seenivasagam, "Review of Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques", ICTACT Journal on Soft Computing, vol.3, pp. 605–609, July 2013.
[2] Jyoti Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications, vol.17, pp. 43–48, Mar. 2011.
[3] S. Vijayarani and S. Sudha, "An Efficient Clustering Algorithm for Predicting Diseases from Hemogram Blood Test Samples", Indian Journal of Science and Technology, vol.8, pp. 1–8, Aug. 2015.
[4] B.Venkatalakshmi, M.V Shivsankar, "Heart Disease Diagnosis Using Predictive Data Mining", International Journal of Innovative Research in Science, Engineering and Technology, vol.3, pp. 1873–1877, Mar. 2014.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)