A Comparative Study of heart disease prediction USING DATA MINING TECHNIQUES

C.SOWMIYA1*, Dr.P. SUMITHRA2*

Ph.D Research Scholar, Department of Computer Science, Vivekananda College of Arts and Sciences for Woman (Autonomous),Elayampalayam.1

Professor, Department of Computer Science, Vivekananda College of Arts and Sciences for Woman (Autonomous), Elayampalayam.2

ABSTRACT: Heart disease is the number one problem of the world. Heart disease more than people deaths occur during the first heart attack .But not only for heart attack have some problems attacked for breast cancer, lung cancer, ventricle. Valve, etc… a common heart disease is nothing .but a cardiovascular disease or a coronary heart disease is a very dangers disease.most people attack for heart disease from the world. Coronary heart disease blood vessels around it. This disease causes disability as damage to brain so resulting in death. Data mining techniques are used is due to extract value from very large.In this paper focus on classification techniques was evaluated of prediction of heart disease. Using medical profiles such as age, sex, blood pressure, chest pain type, fasting blood sugar. It can predict like of patients getting heart disease.

Keywords: Data mining, Heart disease, classification, cluster,

———————————◆—————————

I.INTRODUCTION

Now day's Heart disease is main reason for death in the world. Heart disease is a number one problem from the world. Heart disease is leading cause of death in the world. Over the past few years. Some types of disease occur from heart. There are several kinds of causes, reason, factor which increase the risk factor of heart disease. There are considered as important reason of heart disease. Most of hospitals admitted in heart disease patient. This disease mostly affected in male because smoking habits. This paper analyzes the different kinds of heart disease using the classification techniques.

II.HEART DISEASE

Heart is important part of our human body. More than country affected for heart disease every year some of the million people for death from heart disease. Life is itself dependent on efficient working a heart.as brain, circulation of blood in body is inefficient the organs like brain suffer and if then heart is not properly within it. More then disease attack for heart. now days many hospitals not proper treatment. but increasing the payment of bill. some of hospitals average treatment for patients so result is better. Now day's use of computer technology in the fields of medicine area diagnosis. Heart disease is a group of condition affecting the structure and function of heart and has more root causes. Heart disease is the leading cause of every year death in the world. Some types of disease occurs attack for heart. Types of disease considered are **coronary heart disease, angina pectoris, congestive heart failure, cardiomypathy, congenital heart disease, arrhythmias, myocarditis, heart attack; heart cancer etc.** in this disease is particularly specific very dangers disease to cardiovascular disease or coronary heart disease. There are considered some important reasons of heart disease.

Age Smoking Sugar Obesity Depression Hyper tension High blood cholesterol poor diet Family history Physical inactivity.

III.DATA MINING

Data mining is huge amount of several data base. Data mining is an intelligent creative process. Data mining is used various fields. Now days data mining many places using.This data mining techniques are many advantages and efficient that can be heart disease prediction. Different types of data mining techniques available are classification, cluster, feature selection, association rule can be analyzing the heart disease prediction. . Data mining tools perform data analysis and may uncover important data patterns, contributing greatly to business strategies, knowledge bases, and scientific, medical research. Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years due to the wide availability of huge amounts of data.

Now day's data mining using of more than important application fields.

> Medical, Insurance, Telecommunication ,Finance ,Utilities, Data service providers, Transport, Consumer good ,IT ,Railway, ship, banking

Classification

Classification is one of the well known problems under data mining. To classify the data/objects into different classes or groups. For example data can be broken down according to topical content file type, average file size, gigabytes, and megabytes. Classification is the process of learning a function that cans data objects to a subset of a given class set. Some types goals of classification, first finding a good general that can predict the class of but far unknown data objects with high accuracy. Second to find a compact and easy Understandable class model for each other classes.

Cluster

Cluster is a group of objects. For example data elements into different groups of similarity between in a single group cluster partitions the data set in to cluster classes. Each and every near object is neighborhood object. There are two goals of cluster. First one is a inter class second is an intra class. Inter class cluster means cluster distance is maximized. Intra cluster means cluster distances are minimized.

Feature selection

Feature selection also known as variable selection attribute selection and variable subset selection in this process of selecting a subset of relevant features for use in model construction.

Association rule

Association rule mining is a very import rule of data mining techniques. Association rule is identifying of association huge data base and their values. In this pattern creative techniques which does not serve to solve classification problems and predict problems. This paper focus on study of exiting heart disease prediction task by a using data mining techniques and different issues in exiting of heart disease prediction.

IV. LITERATURE REVIEW

k.srinivas etal(2011) "presented application of data mining techniques in healthcare prediction of heart attacks". The powerful use of classification various data mining techniques using such as a decision tree,k-nearest neighgor naïve bayes. Huge volume of health care data using in the data mining tool Tanagra used for conducted data analysis for the learning purpose. This paper consider of 3000 instances for training data set it is various type 14 attributes. The instances data set is providing the result of various types of testing to say the accuracy of heart disease. The performance of the classifiers set of value and results are analyzed. So to the attributes dataset is divided into two parts. Using training data set is 70% and testing dataset is 30%. The comparison of some classification algorithms using. This paper considered the best algorithm naive bayes.this model using the naive bayes taken a time to run data for best result achieved.comparision of

various classifications algorithm hence better performance accuracy in naive bayes.

Algorithm	Accuracy	Time Taken
Naïve bayes	52.33%	609 ms
Decision tree	52%	719 ms
k-nearest neighbor	45.67%	1000 ms

A1-Milli N, (2013) "Back Propagation algorithm method. To improve the multilayer neural networks in a supervised manner. Back propagation algorithm based on error correction learning. This algorithm based through the various layer of network. From the forward and backward class. Compare to oher algorithm better result for back propagation algorithm .this algorithm research to compare classification techniques. The author efficiency and deliver high accuracy from the heart disease prediction.
Anuradha,

srinivasaraghavan,vincyjoseph(2016)
"Comparative analysis of accuracy on heart disease prediction using classification method "This paper research method algorithm is naïve bayes and SVM (Support Vector Machine), logistic regression.
Naive bayes : Naïve bayes algorithm is a good tool in medical diagnosis. The classifier processes each attributes probability in a class. Naïve bayes is a simple good and efficient performance in classification due to good accuracy for used in medical diagnosis.

Support Vector Machine: SVM is a one of the classification method in this type used to recognize patters and data in a regression and classification analysis. SVM is two classes classified recognizes and separates. Similar data by finding the best hyper plane that separates all data points of one class from other class. Mathematical functions are involved in SVM.

Its performance magnify with number of attributes.

Logistic Regression: Logistic Regression is mainly used for prediction besides it can also be used calculating the probability of success. Basically logistic regression involves fitting and equation. It is a type of statistical regression analysis method used for approximation and prediction of result of a dependent attributes. **Dependent means** it can take only some set of values for example good, or bad, true, or false, on, or off.

These papers consider various classification algorithm using. But high accuracy of performance of heart disease prediction in SVM algorithm is better result.

Algorithm	Accuracy
Naïve bayes	75%
SVM(Support vector machine)	80%
Logistic Regression	79%

Atulkumarpandey, Prabhat Pandey, KL J aiswal, Ashish Kumarsen (2013) "Data mining clustering techniques in the prediction of heart disease using attribute selection method" this model approach on heart disease prediction using cluster technique. Cluster is a group of objects. Some type of cluster used in this paper. Some various cluster method using in different accuracy. Density based cluster, model based cluster, grid based cluster, portioning cluster some various type used. The author used in medical data set correctly and incorrectly method used to accuracy in better. But attribute selection model used in proper display heart disease patients. Various types of cluster technique compared. Considered the best performance accuracy of cluster algorithm make density based cluster.

Cluster algorithm	Correctly classified instance	In correctly classified instance	Prediction accuracy
COBWED	6	297	1.9802%
EM	247	56	81.5182
FARTGEST FIRST	223	80	73.5974
MAKE DENSITY BASED CLUSTER	247	56	81.5182
SIMPLE K-MEANS CLUSTER	245	58	80.8581

Mai Shouman, Tim Turner, Rob Stocker(2011) "**Using Decision Tree for Diagnosing Heart Disease Patients**". Particular researches have been investigating the application of the various classification techniques. To help health care management with improved accuracy in the diagnosis of heart disease. Andreev a used c4.5 decision tree in the diagnosis of heart disease. Accuracy of 75.73% (Andreev 2006) this paper comparison of 3 classification algorithm c4.5 and j4.8 and bagging algorithm. Considered the best performance algorithm bagging algorithm. When using the UCI repository heart disease data set. Hence Compare to 3 algorithms in higher accuracy of bagging algorithm 81.41% evaluated performance of heart disease prediction.

Performance comparison algorithm

Algorithms	Accuracy
J4.8 Decision tree	78.9%
Bagging Algorithm	81.41%
C4.5	75.73%

CONCLUSION

In the world every day heart disease problem in attack from people. Doctors' good knowledge and heavy treatment not satisfied for patients. But Every year more than people for death. But not good accuracy result. So High accuracy provided from medical field every year death people decrease. The result of the various classification algorithm compared to the naïve bayes are very encouraging. The difference in the accuracy is noticeable. This paper presentation in comparisons of various classification algorithm and cluster feature selection method using. This paper considered the high accuracy and best performance of using Naïve bays 84% using the heart disease prediction.

REFERENCES

1. Han.j.and m.kamber (2006) Data Mining concepts and techniques, Morgan Kaufmann publishers.
2. T.Srinivas etal.(2011) presented application of data mining techniques in health care and prediction of heart attacks international journal on computer science and engineering.vol.2 No.02 pp 250-255.2011.
3. A1-Milli N, (2013) **"**Back Propagation algorithm neural network for prediction of heart disease". Journal of theoretical and applied information technology 56(1).

4. Anuradha srinivasaraghavan,vincyjoseh (2016) **"**Comparative analysis of accuracy on heart disease prediction using classification method international Journal of applied information systems volume 11-No 2. July 2016.

5. Atulkumarpandey, Prabhat Pandey, KL J aiswal, Ashish Kumarsen (2013) **"**Data mining clustering techniques in the prediction of heart disease using attribute selection method" International journal of science and engineering and Technology Research volume 2, issue 10 2013.

6. Mai Shouman, Tim Turner, Rob Stocker(2011) "Using Decision Tree for Diagnosing Heart Disease Patients. Proceedings of the 9-th Australasian Data Mining Conference (AusDM'11), Ballarat, Australia CRPIT Volume 121 - Data Mining and Analytics 2011.

7. Chau, M.; Shin, D.,"A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms". Proceedings of IEEE International Conference on Dependable, Autonomic and Secure Computing 2009, pp.183-187.]
   www.archive.ics.uci.edu/ml/datasets/Heart+Disease

8. Shadab Adam Pattekari and AsmaParveen "Prediction System for Heart Disease Using Naive Bayes", International Journal of Advanced Computer and Mathematical Sciences. Vol 3, Issue 3, 2012, pp 290294.

9. Justin, T., Gajsek, R., Struc, V., and Dobrisek, S.: Comparison of Different Classification Methods for Emotion Recognition. MIPRO 2010, Opatija, Croatia, pp. 700-703, (2010)

10. 11.Mythili T., Dev Mukherji,Nikita padalia, and abhiram Naidu " A Heart Disease prediction model using SVM Decision Trees-Logistic Regression (SDL)",IJCA,vol.68-N0.16April 2013

11. Sudhakan k.2014 Study of Heart disease prediction using data mining 4(1):115.1160.

12. Chaitrali. Danger and sulabha s.Apte, Improved Study of Heart disease prediction system using data mining classification techniques. International journal of computer applications vo1.47, No10, pp, 0975-888, 2012.

13. Rahman Rashedum.and Farhana Afroz.2013 comparison of various classification techniques using different data mining tools to diabetes diagnosis journal of software Engineering and applications.

14. Dilip Roy Chowdhury, Mridula Chatterjee & R. K. Samanta, An Artificial Neural Network Model for Neonatal Disease Diagnosis, International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (2): Issue (3), 2011.

15. Vanisree K, Jyothi Singaraju, Decision Support System for Congenital Heart DiseaseDiagnosis based on Signs and Symptoms using Neural Networks, International Journal of Computer Applications (0975 – 8887) Volume 19– No.6, April 2011.

16. Milan Kumari, Sunila Godara, Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction, IJCST Vol. 2, Iss ue 2, June 2011