Predicting Cardiovascular Diseases Risk using Data Mining Techniques

R Pavithra[1], V Pavithra[2]

PG Scholar, Department of Computer Science, Stella Maris College, Chennai, India[1].

Asst.Professor, Department of Computer Science, Stella Maris College, Chennai, India[2]

Abstract— cardiovascular diseases are the most common disease that causes death all around worldwide. Not only the elder (older) people, people of all ages affected by heart problem diseases. Prediction of the cardiovascular problem is utmost important to save the life of humans. It is essential and important to diagnose and manages the heart disease problem. A large number of people are affected by heart disease problem and hence it is quite difficult to predict adequately. Data mining is the technique helps to collect the data's and also retrieve the hidden data. Data mining techniques provide services for an advanced development in the field of medicine. With the help of Data mining, it is very easy to access the information about the risk factors of heart disease problem for proper diagnosis and treatment. Data mining techniques (algorithms) can save a huge amount of people from mortality. Data mining contains various forms of algorithms which provide an accurate set of results.

Keywords— *Naïve Bayes, Random Forest, Prediction, Data mining, Heart Disease.*

_____*****_____

I. INTRODUCTION

The cardiovascular is a main, truly remarkable organ, pumping over 7500 litres of blood a day and beating more than 40 million times a year. The cardiovascular system is the first organ system to become fully functional in utero. Cardiovascular disease remains the leading contributor to mortality worldwide. As per the medical terms, congestive heart failure (CHF) is a general term for heart failure. It is referred as an endpoint of various forms of cardiac disease which is a progressive condition that carries an extremely poor prognosis. Most cases of heart failure are due to systolic Dysfunction (characteristically a consequence of hypertension) or can occur in normal hearts suddenly burdened with an abnormal load (e.g., with fluid or pressure overload).

Data mining methods and techniques in healthcare support for effective treatment and decision making. Healthcare centers face a major challenge provision of quality services; it implies diagnosing patients and treatments. Data mining can be used to retrieve all kinds of data that gives details about the past and predicting the future by analyzing the data. All kinds of field/organization have stored large amounts of data over years, data mining helps to extract valuable knowledge of data's. Data mining helps in the process of automating information. Data mining techniques have been most widely used in healthcare applications by means of prediction. It plays a major role in medical fields. A huge amount of data can be analyzed in data mining to get accurate results. Not only known data's but also hidden data is too analyzed. Predicting of cardiovascular diseases using data mining techniques has found a good solution.

To extract patterns from the huge amount of data, powerful techniques of data mining can be used. It automatically analyses, classify and summarize the data into useful information. The database can also be used to extract information, but it can only retrieve only the past (happened) information, whereas data mining can be used to access the old record as well as to predict the future.

Feature No.	Feature Name
1	age
2	sex
3	chest
4	resting_blood_pressure
5	serum_cholestoral
6	fasting_blood_sugar
7	resting_electrocardiographic_results
8	maximum_heart_rate_achieved
9	exercise_induced_angina
10	oldpeak
11	slope
12	number_of_major_vessels
13	thal
14	class

Table:1:Feature in the Dataset

II. LITERATURE SURVEY

In this[1] paper the researcher has used Artificial Neural Network (ANN) for predicting Heart's stroke, this uses Back-propagation algorithms to train and test various categories of stroke diseases by ANN architecture. The data are generated in a standardized way so that is free from error. Finally, data's are analyzed from data set to define its anomalies and column parameters.

The [2] main purpose of this paper is to build an intelligent heart prediction system by extracting hidden information from

the huge amount of data's. The researcher has used Multilayer Perceptron Neural Networks (MLPNN) which helps in mapping input data set onto an appropriate layer which is used to develop an intelligent way diagnosing heart diseases

This paper [3] proposes how to predict cardiovascular diseases using data mining, it uses a Naïve Bayes algorithm to find the frequently appeared attributes of the prediction model, which is then Weighted Association Rule is applied to generate the interesting pattern.

This paper is [4] mainly developed to predict heart diseases using Machine Learning Algorithms, where the performance accuracy of data mining algorithms such as Naïve base classifier, Support Vector Machine, Decision Tree, K-Nearest Neighbour have been compared.

This paper [5] analyzed a heart disease data set using the Naïve Bayes classifier. This paper specifies the features of each category. Here web application is developed that allows users to get instant guidance on their heart disease, where the web application Include various details about the heart disease.

In this paper [6] the researcher uses Naïve Bayes' classifier to predict hypertension as it is the main problem that our society faces. The researcher considered a data set which consists of identified variables for diagnosing hypertension to develop a predictive model. This paper also evaluates the performance of the Naïve Bayes algorithm by using WEKA tool.

III. PROPOSED METHOD

The algorithm used for predicting the risk of heart diseases are Naïve Bayes and Random Forest, they are the machines learning (data mining) technique which is mainly applicable when the input data sets are huge in number. Naïvebayes is an old method for predictor selection and classification. Naïvebayes algorithm is based on Bayes Theorem. It provides a model which is capable of predicting and describing of patients with heart diseases. Naïve Bayes produces the highest adequate solutions to save enormous peoples from cardiovascular problems. In terms of data mining, the random forest algorithm is known a random forest classifier. It can be used for both classification and regression to solve problems. As the name implies it is a forest made up of a larger number of trees. There is a correlation between the number of trees and the accuracy of random forest, higher the number of trees increases the accuracy results of random forest. It works efficiently even on a large amount of data's. Hence, it is used to produce higher accurate solutions.

The proposed approach is explained as follows,

Step 1: Dataset is loaded into R Tool.
Step 2: Null observations are then removed using R function.
Step 3: Dataset is then classified using Naïve Bayes technique and obtained a result.
Step 4: The outcome of the previous step is applied as an input to classify the Dataset using Random Forest algorithm.
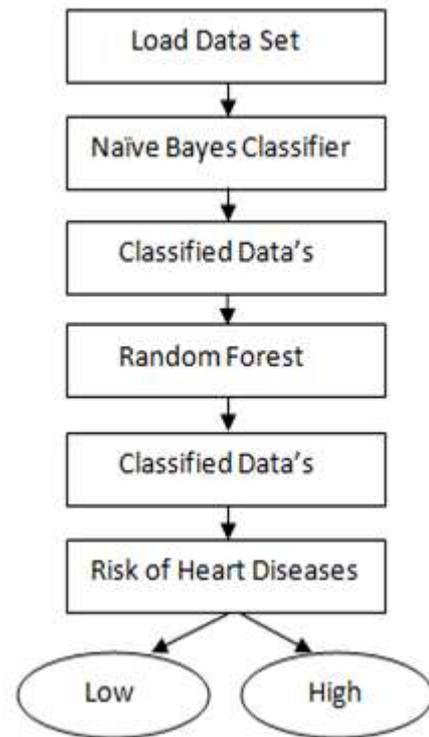**Step 5:** Stop

Fig: 1: The Proposed Scheme

The proposed approach uses a Naïve Bayes and Random Forest technique to get an adequate result. This increases the accuracy of predicting data's. The work is implemented using R environment.

R environment is a programming language and it provides the huge variety of statistical analyses such as classification, clustering, series analysis. It helps handling data effectively, storage and graphical facilities. R is available as Free Software which is the main advantage of the R environment.

IV. EXPERIMENTAL RESULTS

This paper proposes the risk prediction of cardiovascular disease using data mining techniques. The proposed scheme is executed in R Software Environment. The implemented results obtained are described in this section. The major criteria of machines learning (data mining) meant to be its Accuracy level. To prove the effectiveness of the proposed system, it should obtain the major criteria called accuracy. This paper has been implemented the data set using Naïve Bayes technique and Random Forest techniques, which classifies the data set's and predict the risk of cardiovascular disease with a higher accuracy of 94%.
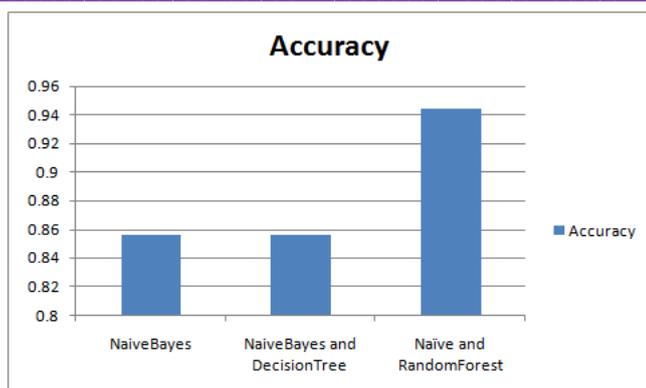
V. CONCLUSION

A detailed elaboration about the techniques, role, and importance's of data mining in a predicting the risk of heart disease has been presented. Thus, we come to know that data mining plays an important role in the field of medicine due to its effective diagnosis. Because of its accuracy in adequacy data, it has stepped into all fields. The data mining algorithm also is used in various fields with results in great adequate results.

REFERENCE

[1] D. Shanthi, D., Sahoo, G., & Saravanan, N. (2009). Designing an artificial neural network model for the prediction of thrombo-embolic stroke. *International Journals of Biometric and Bioinformatics (IJBB)*, *3*(1), 10-18.

[2] Chaitrali S. Dangare and Sulabha S.Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications, Vol.47, No. 10, pp.0975-888, 2012

[3] Patel, A., Gandhi, S., Shetty, S., & Tekwani, B. (2017). Heart Disease Prediction Using Data Mining. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 - 0056 Volume: 04 Issue: 01 | Jan -2017

[4] Sen, S. K. (2017). Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms. *International Journal Of Engineering And Computer Science (IJECS)*, *6*(6), 21623-21531.

[5] Garima S, Kiran B, Shivani S, Shraddha S, Sulochana D. Heart disease prediction using Naïve Bayes. International Research Journal of Engineering and Technology (IRJET). 2017;04(03).

[6] Afeni, B. O., Aruleba, T. I., & Oloyede, I. A. Hypertension Prediction System Using Naive Bayes Classifier., Journal of Advances in Mathematics and Computer Science 24(2): 1-11, 2017; Article no.JAMCS.35610.