Prediction of Heart Disease using Classification Algorithms

M.Lavanya[1], Mrs.P.M.Gomathi[2],
Research scholar, Head of the department,

Department of computer science,

P.K.R college for women, gobi

Abstract- Data mining is an iterative progress in which evolution is defined by detection, through usual or manual methods. The discovered knowledge can be used for different applications for example healthcare industry. The heart disease accounts to be the leading cause of death worldwide. It is difficult for medical practitioners to predict the heart attack as it is complex task that requires experience and knowledge. Different data mining techniques such as association rule mining, classification, clustering are used to predict the heart disease in health care industry. data mining algorithm such as J48,naïve bayes, REPTREE, Neural networks, CART are applied in this research for predicting heart attacks.

Keywords: heart disease, J48, naïve bayes, REPTREE, Neural networks, CART.

I.INTRODUCTION

Data mining is process of extracting useful information from large amount of database. Data mining is most useful in an exploratory analysis because of non trivial in large volumes of data. Predictions and descriptions are principal goals of data mining ,in practice. Prediction in data mining involves attributes or variables in the data set to find an unknown or future state valus of other attributes.

II. HEART DISEASE

The heart is important organ of our body. Life is dependent on efficient working of heart . if operation of heart is not proper, it will affect the other body parts of human such as brain, kidney etc.

The features of increase the possibility of heart attacks are smoking, lack of physical exercise, high blood pressure, high cholesterol, unhealthy diet, harmful use of alcohol, and high sugar levels, cardio vascular disease ,hypertension heart, congenital heart.

III.RELATED WORKS

The data mining techniques includes different works to explore a variety of disease such as cancer, diabetes, heart disease. Heart disease is the most important reason of fatality in the south Africa, Canada, and England. Hearts disease kills individual in each 32 seconds in the world. Jyoti soni et al proposed three different supervised machine learning algorithms for

2173

heart disease prediction. Classification is supervised learning methods to extract models relating main classes of data. Decision tree, naïve bayes and classification via clustering are the three classifiers used to analyze the occurrence of heart disease for the patients

IV. PATIENT DATASET

The patient data set is complied from data collected from medical practitioners in south Africa. Only 11 attributes from the database are considered for the predictions required for the heart disease.

The following attributes with nominal values are considered : patient identification number, gender, cardiogram, age , chest pain, blood pressure level, heart rate, cholesterol, smoking ,alcohol consumption and blood sugar level.

Waikato environment for knowledge analysis has been used for prediction due its proficiency in discovering, analysis and predicting patterns.

V. DATAMINING TECHNIQUES USED FOR PREDICTIONS

5.1. decision trees:

The decision tree approach is more powerful for classification problems. There are two steps in this techniques building a tree and applying the tree to the dataset. There are many popular decision tree algotithms CART,ID3, C4.5,CHAID, and J48. From these J48 algorithm is used for this system. J48 algorithm uses pruning method to bulid a tree. Pruning is a technique that reduces size of tree by

removing over fitting data, which leads to poor accuracy in predictions. The J48 algorithm recursively classifies data until it has been categorized as perfectly as possible. This techniques gives maximum accuracy on training data.

5.2. naïve bayes:

Naïve bayes classifier is based on bayes theorem. The classifier algorithm uses conditional independence, means it assumes that an attribute value on a given class is independent of the values of the other attributes.

The bayes theorem is follows:

Let X={X1,X2,…..xn) be a set of n attributes.

X is considered as evidence and H is some hypothesis, the data of X belongs to specific class C.

5.3.neural networks:

An artificial neural network often just called a "neural network" is the mathematical model or computational model based in biological neural network. In is an emulation of biological neural system .

It maps a set of input data onto a set of appropriate output data. It consists of 3 layers input layer, hidden layer, output layer. The primary function of neurons of input layer is to divide input xi into neurons in hidden layer neuron of hidden layer adds input signal xi with weights wji of respective connections from input layer. The output function Yj is function of yj=f(E wji xi)where f is a simple thresholds function

2174

such as sigmoid or hyperbolic tangent function.

VI. EXPRIMENTAL RESULTS

They are the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSR) [10]. The mean absolute error (MAE) is defined as the quantity used to measure how close predictions or forecasts are to the eventual outcomes. The root mean square error (RMSE) is defined as frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed.

Techniques	Correctly classified (%)	Incorrectly (%)	MAE	RMSE	RAE (%)	RRSE (%)
Naïve Bayes	85.92	14.07	0.169	0.33	34.22	68.18
IBK(KNN)	100	0	0.00	0.00	0.90	0.73
J48(desion tree)	91.85	8.14	0.14	0.26	28.59	134.2
CART	95.92	4.07	0.07	0.19	14.73	38.38
ANN	99.25	0.74	0.02	0.09	4.50	18.29
SMO	85.55	14.44	0.14	0.38	29.24	76.48

VI .CONCLUSION

Medical related information are huge in nature and it can be derived from different birthplaces which are not entirely applicable in feature. The research undertook an experience on application of varies data mining algorithm to predict the heart attacks and to compare the based method of prediction. The predictive accuracy determined by J48,REPTREE, naïve bayes, neural networks, CART. The overall objective is to study the various data mining techniques available to predict the heart disease and to compare them to find the best method of prediction.

REFERENCES

1). Sellappanpalaniappan, rafiah awing, "*intelligent heart disease prediction system using data mining techniques*", IJCSNS international journal of computer science and network security, vol.8 ,aug 2008.

2). V. manikantan and S. latha ,"*predicting the analysis of heart disease symptoms using medical data mining methods*", international journal of advanced computer theory and engineering, vol 2,2013.

3). A.K. sen,S.B. patel, and D.P .shukla," *A data mining techniques for predicting of coronary heart disease using neuro-fuzzy integrated approach two level*", international journal of engineering and computer science, vol 2, vol 1,2013.

4). G. Subbalakshmi, K. ramesh and N.chinna rao "*decision support in heart disease prediction system using naïve bayes*",issn 0976-5166, vol2,2011.

5).N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra, *performance analysis of classification data mining techniques over heart diseases data base*, international journal of engineering science and advanced technology, 2012.