International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 01 | Jan -2017 www.irjet.net p-ISSN: 2395-0072

Heart Disease Prediction Using Data Mining

Ajad Patel, Sonali Gandhi, Swetha Shetty,Prof. Bhanu Tekwani

[1]*Ajad Patel, Dept. Of Information Technology, Vidyalankar Institute Of Technology, Maharashtra, India*
[2]*Sonali Gandhi, Dept. Of Information Technology, Vidyalankar Institute Of Technology, Maharashtra, India*
[3]*Swetha Shetty*, Dept. Of Information Technology, Vidyalankar Institute Of Technology, Maharashtra, India
[4] *Prof.Bhanu Tekwani*, Dept. Of Information Technology, Vidyalankar Institute Of Technology, Maharashtra, India

---

Abstract – *Heart disease is a major life threatening disease that cause to death and it has a serious long term disability. There is wealth of data available within the health care system. However, there is lack of effective tools to discover hidden relationships and trends in data .Advanced data mining techniques can help remedial situations. This paper describes about a prototype using data mining techniques mainly Naïve Bayes and WAC (Weighted Associated Classifier).*
The dataset is composed of important factors such as age ,sex, diabetic, height, weight, blood pressure, cholesterol, fasting blood sugar, hypertension, disease. The system indicates whether patient had a risk of heart disease or not.

***Key Words*: Data mining, Naïve Bayes, WAC, Prediction**

1. INTRODUCTION

It is a world known fact that heart is the most essential organ in human body if that organ gets affected then it also affects the other vital parts of the body. Data mining aids in healthcare to support for effective treatment, healthcare management, customer relation management, fraud and abuse detection and decision making. A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems.

The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making. Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. For instance it might now be possible for the physicians to compare diagnostic information of various patients with identical conditions. Likewise, physicians can also confirm their findings with the conformity of other physicians dealing with an identical case from all over the world. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial.

1.1 Data Mining

Data Mining is about explaining the past and predicting the future by means of data analysis. Data mining is a multi-disciplinary multi-disciplinary field which combines statistics, machine learning, artificial intelligence and database technology. The value of data mining applications is often estimated to be very high. Many businesses have stored large amounts of data over years of operation, and data mining is able to extract very valuable knowledge from this data. The businesses are then able to leverage the extracted knowledge into more clients more sales, and greater profits. This is also true in the engineering and medical fields. Data mining predicts the future of modeling.

Predictive modelling is a process by which a model is created to predict an outcome. If the outcome is categorical it is called categorical and if the outcome is numerical it is called regression. Descriptive modeling or clustering is assignment of observations into clusters so the observation of same cluster are similar

1.2 Heart disease in India

Cardiovascular diseases (CVDs) have now become the leading cause of mortality in India. A quarter of all mortality is attributable to CVD. Ischemic heart disease and stroke are the predominant causes and are responsible for >80% of CVD deaths. The Global Burden of Disease study estimate of age-standardized CVD death rate of 272 per 100 000

population in India is higher than the global average of 235 per 100 000 population. Premature mortality in terms of years of life lost because of CVD in India increased by 59%, from 23.2 million (1990) to 37 million (2010).

2. Naïve Bayes

The Naive Bayes model is an old method for classification and predictor selection that is enjoying a renaissance because of its simplicity and stability.
The following notation is used:

J_0	Total number of predictors.
X	Categorical predictor vector
X'	(X1, ..., XJ), where J is the number of predictors considered.
M_j	Number of categories for predictor Xj.
Y	Categorical target variable.
K	Number of categories of Y.
N	Total number of cases or patterns in the training data.
N_k	The number of cases with Y= k in the training data.
$N^j mk$	The number of cases with Y= k and Xj=m in the training data.
πk	The probability for Y= k.
$p^j mk$	The probability of Xj=m given Y= k.

The Naive Bayes model is based on the conditional independence model of each predictor given the target class. The Bayesian principle is to assign a case to the class that has the largest posterior probability. By Bayes' theorem, the posterior probability of Y given X is:

Let X1, ..., XJ be the J predictors considered in the model. The Naive Bayes model assumes that X1, ..., XJ are conditionally independent given the target; that is:

These probabilities are estimated from training data by the following equations:

2. WAC

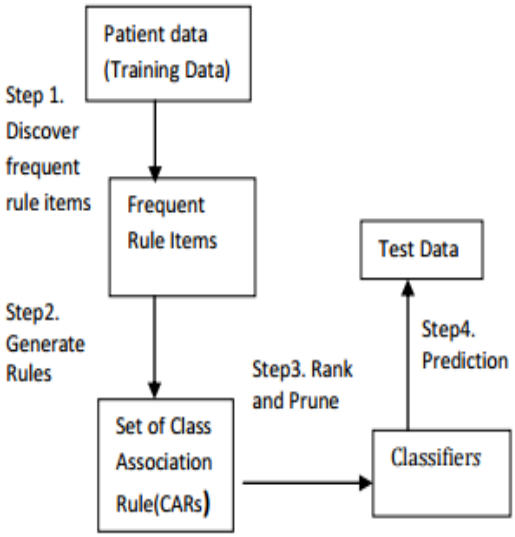

Figure -1: Associative Classifier for Data Mining

Weighted Associative Classifier (WAC) is a new concept that uses Weighted Association Rule for classification. Weighted ARM uses Weighted Support and Confidence Framework to extract Association rule from data repository. The WAC has been proposed as a new Technique to get the significant rule instead of flooded with insignificant relation.

The major steps are as follows:

1: Initially, the heart disease data warehouse is pre - processed in order to make it suitable for the mining process.

2: Each attribute is assigned a weight ranging from 0 to 1 to reflect their importance in prediction model. Attributes that have more impact will be assigned a high weight (nearly 0.9)and attributes having less impact are assigned low weight(nearly 0.1) .

3: Once the preprocessing gets over, Weighted Association Rule Mining (WARM) algorithm is applied to generate interesting pattern. This algorithm uses the concept of Weighted Support and Confidence framework instead of tradition support and

confidence. Rules generated in this step are known as CAR (Classification Association Rule) and is represented as X ◊ Class label where X is set of symptoms for the disease. Example of such rules are (Hypertension, "yes") ◊Heart_Disease="yes" and {(Age," >62"), (Smoking_habits, "yes"), (Hypertension, "yes")}◊Heart_Disease="yes" .

4: These rules will be stored in Rule Base.

5: Whenever a new patient's record is provide, the CAR rule from the rule base is used to predict the class label.

3. CONCLUSIONS

In this paper, we have presented an intelligent and effective heart disease prediction methods using data mining. We studied an efficient approach for the extraction of significant patterns from the heart disease data warehouses for the efficient prediction of heart disease. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial. Data mining have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

The proposed work can be further enhanced and expanded for the automation of Heart disease prediction. Real data from Health care organizations and agencies needs to be collected and all the available techniques will be compared for the optimum accuracy.

REFERENCES

[1] C. S. Dangare, S S. Apte Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques , International Journal of Computer Applications (0975 – 888) Volume 47– No.10, June 2012.

[2] J . Soni, U. Ansari, D. Sharma, S. Soni. Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications (0975 – 8887)Volume 17– No.8, March 2011

[3] K. Srinivas, B. Kavihta Rani, Dr A. Govrdhan. Application of data mining techniques in healthcare and prediction of heart attacks

[4] N .Aditya Sundar, P. Pushpa Latha, M. Rama Chandra, Performance analysis of classification data mining techniques over heart disease database, IJESAT volume-2, Issue-3, 470 – 478

[5] Nilakshi P. Waghulde, Nilima P.Patil, Genetic Neural Approach for heart disease prediction, International Journal of Advanced Computer Research, Volume-4 Number-3 Issue-16 September-2014