International Journal of Engineering Sciences &Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164

IJESRT

RESEARCHERID
THOMSON REUTERS

[Chandrasekhar * *et al.*, 8(3): March, 2019]
IC™ Value: 3.00

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

DETECTION OF HEART DISEASES USING DATA MINING TECHNIQUES

T.Chandrasekhar[*1], S.S.V Harsha Pavan[2] & B.Vinay[3]

[*1,2&3]Department OF Computer Science Engineering, Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam

DOI: 10.5281/zenodo.2608158

ABSTRACT

Data mining is the very apt technology, which includes the process of mining actionable information from huge dataset, which is used to analyze large masses of data and derive patterns or statistics that can be converted to knowledgeable implementation. Medical data mining has a great potential for deriving the hidden patterns in the data sets of medical industry. The derived patterns can be utilized to do clinical diagnosis. These data need to be collected in a standardized form. From the medical profiles fourteen attributes are extracted such as age, sex, blood pressure and blood sugar etc. can predict the likelihood of patient getting heart disease. These attributes are fed in to ID3 Algorithm and CART algorithm or Decision tree classification in heart disease prediction, applying the data mining technique to heart disease prediction; it could give a reliable performance that might result in diagnosing heart disease. By this medical industries could offer better diagnosis and treatment of the patient to attain a good quality of services. The main advantages of this paper are: early detection of heart disease ,providing measures for diagnosis and proving the efficiencies of ID3 and CART algorithms.

KEYWORDS: Heart disease; Data mining; Decision tree; ID3 Algorithm,CART Algorithm.

1. INTRODUCTION

In the modern lifestyle health diseases are increasing tremendously. Our life style had a great impact on our health causing heart diseases and other health problems. Taking a survey of present population it is seen that about sixty percentages are suffering from heart diseases.

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.This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.

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.The Global Burden of Disease study estimate of age standardized CVD death rate of 272 per 100,000

In most cases it is noticed at the final stages of disease or after death. The cost of treatment for heart disease is very expensive. The treatment cost is not affordable for everyone. Therefore people are reluctant to do proper treatment at early stages of disease. The aim of our project is to predict the disease at early stage at affordable cost. By using data mining technique we can detect disease at early stage and this will helps to cure the disease by proper diagnosis.

Basic Data Mining techniques are
1.Association
2.Clustering
3.Classification

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

The Database containing huge amount of data with hidden information that used for making decisions. Classification model use to extract a model describing important classes. Classification techniques are ID3 algorithm generating decision trees and Naïve Bayes algorithm . Decision tree are very flexible, easy to understand and easy to debug. It takes care of various issues like missing value, outlier and identifying significant dimensions. Naïve Bayes is supervised algorithm.it assume underlying probabilistic model. It is assumption, so loss accuracy and if no occurrence of attribute or class label then probability estimate will be zero. So decision tree is better than naïve Bayes. Here we used only Classification method for effective detection.

The simple k-means approach it is first difficult to **find the correct number** of clusters you want.You could run a different algorithm first to see the distance between any number of clusters,,repeated centroids may result in increased time complexity.Reverse engineering process should be made to retrieve the result if clustering and association techniques are implemented.

2. RELATED WORKS

- The researchers have been investigating the use of data mining techniques to detect heart disease. There are some factors such as factors associated with heart disease like age, sex, chest pain, blood pressure, cholesterol, blood sugar, etc. These factors are used to diagnosis the heart disease in patients. The researcher intends to provide a survey of current techniques of data extraction from databases using data mining techniques that are used in Heart Disease Prediction.The techniques used here are Naive Bayes, Decision List and KNN. Here the Classification based on clustering is not performing well.

- P.K Anooj et al. presented a weighted fuzzy rule-based system for the diagnosis of heart disease, the system will automatically retrieve knowledge from the patient's data. The proposed system for the prediction of heart disease consists of two phases: (1) automated approach for the generation of weighted fuzzy rules and (2) developing a fuzzy rule-based decision support system. The weighted fuzzy rules were used to build the system using Mamdani fuzzy inference system.

- Nidhi Bhatla et al. [3] proposed to analyse various data mining techniques used in heart disease prediction. The observations reveal that neural networks with 15 attributes has outperformed over all other data mining techniques. Another conclusion from the analysis is that decision tree has also shown good accuracy with the help of genetic algorithm and feature subset selection .

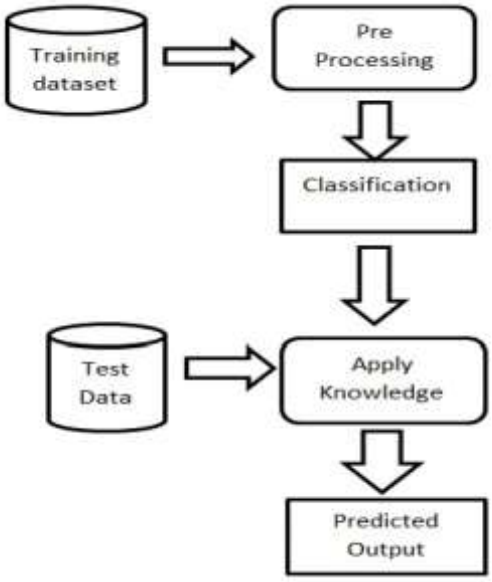
- Aditya Methaila et al. desire to use data mining Classification Modeling Techniques, such as Decision Trees, Naïve Bayes and Neural Network, in addition to weighted association Apriori algorithm and MAFIA algorithm in Heart Disease Prediction .

- Shimpy Goyal et al. discussed Data Mining Techniques to Predict Heart Disease based on K-means and apriori algorithm. The researchers also presented the challenges in detecting and diagnose the diseases and analyze results of research.

3. PROPOSED SYSTEM

The proposed system predicts the heart diseases at an early stage accurately using data mining techniques.A huge amount of healthcare data,which are not mined to discover hidden info for effective decision making. The proposed system structure explains modules of the application and how data is being mined.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7



Genetic algorithm

A genetic algorithm (GA) is a searching that imitate the process of natural evolution. These prerequisites is routinely used to generate useful solutions to optimization and search problems. In our system the genetic algorithm is used to extract attribute from a hugh attribute set.

The extracted attribute are as follows,

1. age: age in years
2. sex: sex (1 = male; 0 = female)
3. cp: chest pain type

Value 1: typical angina
Value 2: atypical angina
Value 3: non-anginal pain
Value 4: asymptomatic

4. trestbps: resting blood pressure (in mm Hg on admission to the hospital)
5. chol: serum cholesterol in mg/dl
6. fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. restecg: resting electrocardiographic results

Value 0: normal
Value 1: having ST-T wave abnormality (T wave inversions a and/or ST elevation or depression of > 0.05 mV)
Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

8. thalach: maximum heart rate achieved
9. exang: exercise induced angina (1 = yes; 0 = no)
10. oldpeak = ST depression induced by exercise relative to rest 1 11. slope: the slope of the peak exercise ST segment

Value 1: upsloping
Value 2: flat
Value 3: downsloping

12. ca: number of major vessels (0-3) colored by fluorosible 3. Thal:
13 = normal; 6 = fixed defect; 7 = reversible defect
14. Num: diagnosis of heart disease (angiographic disease status) Value 0: < 50% diameter narrowing
Value 1: > 50% diameter narrowing

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Algorithm 1: ID3 Algorithm

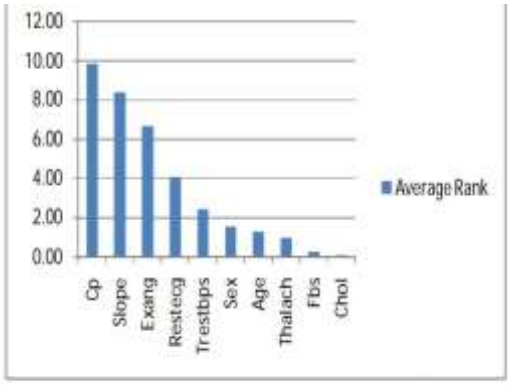
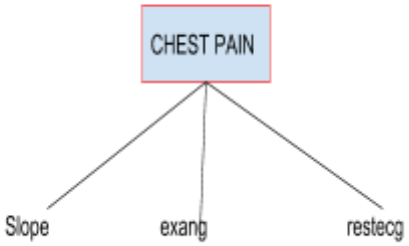


Figure 5: Comparison between importance of attributes

Decision tree algorithms convert raw data to rule or condition based decision making trees. Here, ID3 is one of the common decision tree algorithm. It was introduced in 1986 and it is acronym of **Iterative Dichotomiser**. dichotomisation stands for dividing into two completely opposite pats.So, the algorithm iteratively divides attributes into two sections which are the most dominant attribute and others to build a tree. Thereafter the entropy and information gains of each attribute are calculated. In this way, the most dominant attribute is found. Followed by, the most dominant one is put on the tree as decision node. Then the, entropy and gain scores would be calculated again among the other attributes. Thus, the next most dominant attribute is found. Finally, this procedure processes until we reach a decision for that branch. That's why, it is called Iterative Dichotomiser.We can summarize the ID3 algorithm as illustrated below

$$Entropy(S) = \sum - p(I) \cdot \log_2 p(I)$$

$$Gain(S, A) = Entropy(S) - \sum [p(S|A) \cdot Entropy(S|A)]$$



Algorithm 2: CART Algorithm

Classification and Regression Trees implies the technique of recursively separating the observations in branches to build a tree to improve the prediction accuracy and predicts continuous dependent variables and categorical predictor variables.The CART algorithm was popularized by Breiman et al. (Breiman, Friedman, Olshen, & Stone, 1984; see also Ripley, 1996). Although after many investigations and enhancements by the researchers less research is done on enhancing CART performance in disease diagnosis especially in diagnosis of heart disease. In this paper, a method which is existing is applied to detect heart disease to obtain the result.CART utilizes Gini index to scale the impurity of a section or collection of training tuples. It is capable to handle high dimensional categorical data. Decision Trees can also handle continuous data but they must be turned into categorical data. This simplicity is useful not only for purposes of rapid classification of new observations (it is

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

easier to evaluate one or two logical conditions, than to compute classification marks for possible groups, or predicted values, basing on all predictors and using possibly some complex nonlinear model equations),can also result a simpler "model" to explain why observations are classified or predicted in a particular manner. The final outputs of using tree methods for classification or regression can be summarized in a sequence of logical if-then conditions. Therefore, there is no implicit assumption that the underneath relationships between the predictor variables and the dependent variable are linear,follow specific non-linear link function, or that they are even.

4. CONCLUSION

This paper aims on using different algorithms in data mining and series of several attributes for effective heart disease prediction. Decision Trees have good efficiency using fourteen attributes, and applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute accurate for heart disease prediction. Thereafter, the efficiencies between CART and ID3 algorithm are computed proving that CART algorithm is more efficient than ID3 theoretically and practically.

REFERENCES

[1] Jyoti 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.
[2] P.K. Anooj, Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules, Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40.
[3] Nidhi Bhatla and Kiran Jyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October - 2012 ISSN: 2278-0181.
[4] Aditya Methaila, Prince Kansal, Himanshu Arya, Pankaj.