International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 4 | ISSN : 2456-3307

Analysis and Prediction of Heart Disease Using Decision Tree and Naive Bayes

[1]Megha Rani Raigonda, [2]Vaishnavi

[1]Assistant Professor, Department of MCA, VTUP.G Centre, Kalaburgi, Karnataka, India
[2]P.G Scholar, Department of MCA, VTU P.G Centre, Kalaburgi, Karnataka, India

ABSTRACT

The extraction of the hidden data from the large databases is data mining and it is also known as Knowledge Discovery Mining. It has many tasks one of them used here is predictive tasks which uses, that some variables to predict unknown or future values of other dataset. The major health problem affects large number of people is Heart disease . Unless it is treated at an early stage it causes death. Healthcare industry generates today large amount of complex data about the patients and resources of the hospitals, from the time where there has no sufficient focus on effective analysis tools to discover relationships in data especially in medical sector. The techniques of mining data are used to analyse rich collections of data from different perspectives and derive useful information, to develop diagnosis and predicting system for heart disease based on predictive mining. Number of trials are taken up to compare the performances of different techniques of data mining including Decision tree and Naïve Bayes algorithms. As proposed, the risk factors are taken into consideration ,Decision tree and Naïve Bayes are applied and performance on their diagnosis have been compared by the UCI Machine Learning Repository i,e WEKA tool. Thus, the Naïve Bayes outperforms the Decision tree .

Keywords: Watermarking, Encryption, Distortion, Reliability, Biomedicalimaging,Biomedicaltrasonics, Cryptography, Medical Image Processing, Security of Data

I. INTRODUCTION

Data Mining is exploring a large data sets to pull out in the past unfamiliar data, associations that are hard to notice with numerical methods. It has very importance for providing diagnosis and understanding of medical data. It is very important step in discovering from databases and comprises an iterative sequence and accomplished using prediction, classification, association and many more series of analysis. Although applying data mining in diagnosing in disease is beneficial which can improve the performance and provides reliable accuracy. It is implemented as web based application. The objective of the current try is to discover the aspect of of health care by machine learning procedures, to suggest the automated dig nosing is taken into consideration. Health organization reduce cost ,accomplishing of computer based data . The services that incorporate patients records and there are many tasks of mining the data they are namely predictive and descriptive tasks which consists of many branches were predictive tasks consists of classification, Prediction and time series analysis .were as the descriptive analysis consists of associations of the data clustering the data and then the last is the summarization of the data sets present. Here we consider the predictive analysis of mining the data were we predict predictive analysis process has many steps which namely mentioned in the above figure Were first there is a need of the defining a project is the first step and secondly there is a need of collection of the data ,next the analysis of the data ,next the statistics is produced based on the analysis and later it is modelled and lastly it is deployed these are the various steps to analysis of the Predictive Process.

The recognition of heart disease from a variety of factor or indication is a multi-layered issue which is not free from false guess often go together with by unfixed effects. Thus the try to construct use of information and understanding of frequent specialists and clinical screening data of patients collected in databases to

facilitate the diagnosis process is considered a valuable option. as long as valuable services at affordable costs is a major constraint encountered by the healthcare organizations . important quality service denotes the accurate diagnosis of patients and providing efficient treatment. Poor clinical decisions may lead to disasters and hence are seldom entertain. Besides, it is essential that the hospitals decrease the cost of clinical test. Appropriate computer-based information and/or decision support systems can aid in achieving clinical tests at a reduced cost. Heart is vital part of body and it is caused due to blocking of coronary arties which is caused by deposition of fat on wall of inner artery an also due to cholesterol build .Identifying the major risks from patients of heart to identify the level of risk factors in helping patients to save their life in advance.

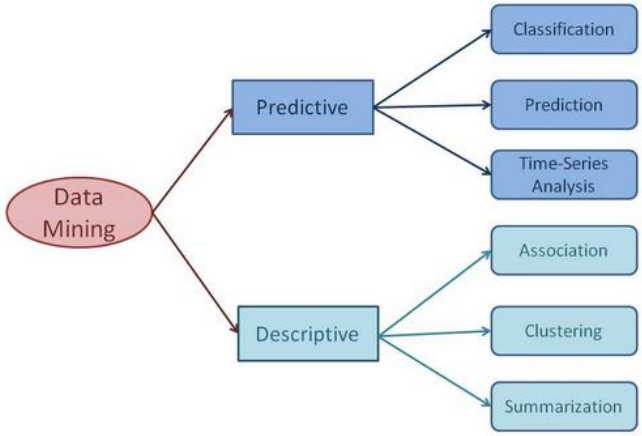


Figure :1 General architecture of datamining tasks

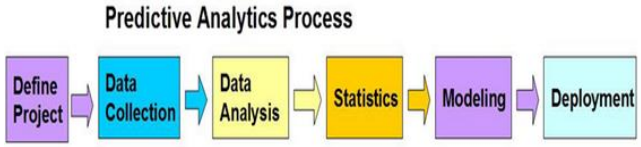


Figure :2 General Predictive analysis Process

PROJECT DESCRIPTION

Mining is the most vital and area with finding out meaningful information from huge data sets. It has become more popular in health care sector as there is a need of proficient and logical meth0d0logy for detecting unfamiliar valuable information from health sector. Tools of data mining performs analysis of data and also uncover patterns. It is more convenient tool to assist physicians in detecting the disease by obtaining information regarding patients data set. There is a need of medical pardoners to predict before it appear or occur in patients. As the huge amount of information is produced in medical associations yet this is not useful ,and there is an absence of successful analysis methods to connections and patterns in healthy care information.

PROBLEM STATEMENT

The main cause of death worldwide is considered by heart disease . Medical Practitioners Generated data with a wealth of unseen information present, and not properly being used effectively for prediction. persons die having skilled symptom that were not taken into deliberation.

OBJECTIVE OF THE PROJECT

The main goal of this effort is to guess achievable heart attack from the patient's dataset using techniques of data mining and determine which module gives the top percentage of correct prediction for the diagnose.
The very important objectives are:

1.High accuracy is obtained.
2. To help the health care sector&
3. Rate of death is reduced.

SCOPE OF STUDY

The scope of the project is that the health be concerned industry gathers huge amount of statistics and made effectively mined to discover hidden information

II. LITERATURE SURVEY

Frequent studies have been done that have fixed the guess of heart disease. They have used different data mining tools and applied different techniques for diagnosis and achieved different results

In [1] Identifying the major risk factors of Heart Disease categorising the risk factors in an order which causes damages to the heart such as high blood cholesterol, diabetes, smoking, poor diet, obesity, hyper tension, stress, etc. Data mining function and technique are used to spot the point of danger factor to help the patients in taking defence in proceed to save their life.

In [2] explain their paper that conclusion bear inn sympathy sickness prophecy organization is residential using both Naive Bayesian Classification and Jelineck-mercer smothing modus operandi. The arrangement extract veiled awareness from a chronological sensitivity disease database. Jelineck-mercer smoothing

technique is supplementary ineffectual than naiive bayes to expect patients with kindness disease.

into year 2010, a study was conducted for analytical reproduction for the Ischemic Heart Disease (I.HD); they applied sponsor-dissemination neural set of connections (BP,NN), the Bayesian neural network (BN,N), the probabilistic neural connections (P'NN) and support vector contraption (SV'M) to extend classification models for identify IH'D patients on a figures obtain from capacity off cardiac charismatic countryside at place above the torso6. The result shows that BPNN and BNN gave the high cataloguing accuracy of 78.43 %, while RBF pip SVM gave the lowest classification accuracy of 60.78 %. Both polynomial kernel SVM and RBF kernel SVM existing the least and max specificity of 45.45 % and 86.36 %, respectively.

In [3] provide a study of different key attributes that can be used to guess the heart disease. Some attributes are mostly responsible for heart disease. with the purpose of foretell the disease base upon patient's raw indication narrative, different key in parameters can be collected. To help the physicians and health care professionals in the prediction of heart diseases, the add up to of attributes should standard. The system helps the medical practitioners to minimise the time of prediction and improve their practice as well as protect the life of the patient.

In [4] provide a study Heart' Diseases by means of various Data Mining Techniques. When numbers mining technique used separately the accuracy is low. To improve the accuracy value, data mining techniques should be combined together. SVM Classifier with genetic Algorithm contains 95% high accuracy than other technique. In future the techniques are hybrid, the accuracy will high. It will definitely help the patients as well as the medical practitioners to predict the heart disease.

In [5] used the Data Mining algorithm such as J 48, Naïve Bayes, REPTREE, CART, and Bayes Net are practical in this make enquiries for predict heart attacks. The research result shows prediction accuracy of 99%.

In [6] In the year 2012, Nidhi et al used data mining tool Weka 3.6.6 in the analysis of heart disease prediction system and achieved 99.52% using Naive Bayes. She also used TA.NGRA data mining tool but she achieved up to 52.33% by using decision trees. And later used .NET data mining tool and achieved up to 96.5% and used neural networks.

In [7] model a projecting mould for heart disease finding out using data mining techniques from Transthoracic Echocardiography Report dataset that is accomplished off pretty the consistency off feeling disease judgment using echo cardioigraphy.

In [8] the year 2013, Abhishek etal used the data mining tool Weka 3.6.4 in heart disease prediction system. He used J48 technique and achieved 95.56% accuracy and using Naive Bayes he ,achieved 92.42%.

III. METHODS AND MATERIAL

I The' naive Baiyesian classifier, or simple Bayesian classifier, work's as follows:

1. Let s be a tuition dataset of tuple and their correlated cluster of student label. As usual, each tuple, is represent by an n-dimensional quality vector, y=(y1, y2,…, yn), depict n capacity made on the tuple from m attribute, in that order, b1, b2,.., bn.

2. presume that there are m lessons, z1, z2,…, zm. Given a tuple, p, the classifier will guess that p belong to the class have the maximum latter prospect, habituated on p That is, the naïve Bayesian classifier predict that tuple "p" belong to the class Ci if and only if.

$$P(Ci|X) > P(Cj|X) \quad \text{for } 1 \le j \le m, j \ne i$$

Thus we maximize P(Ci|X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis. By Bayes' theorem

$$P(Ci \mid X) = \frac{P(X \mid Ci)P(Ci)}{P(X)}$$

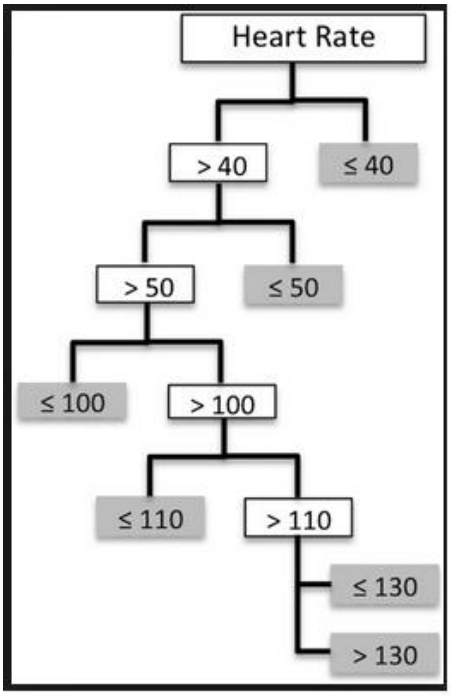
$$P(X \mid Ci) = \prod_{k=1}^{n} P(xk \mid Ci)$$

$$= P(x_1|C_i) \times P(x_2|C_i) \times \dots P(x_m|C_i).$$

Decision Trees

Decision tree learning uses a d (as a predictive model) to go from explanation about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the

extrapolative modelling approach used in statistics, data mining and machine learning. Tree models where the target variable can take a discrete set of values are called **organization trees**; in these ranking structures, leaves represent class labels and branches represent conjunctions of explanation that lead to those class labels. Decision trees where the board inconsistent can take continuous standards (typically real numbers) are called **waning vegetation**.



Decision Tree for prediction of heart disease

A choice tree is a simple description for classify examples. For this section, take for granted that all of the contribution explanation have fixed detached domain, and there is a introverted purpose facade called understanding. each one ingredient of the sphere of the cataloguing is called a division. It arrangement tree is a hierarchy in which each domestic join is considered with an attempt strangeness. The semicircle coming from a swelling labelled with an put in quality are labelled with each of the possible values of the target or productivity attribute or the arc lead to a subordinate conclusion node on a dissimilar input attribute. every piece of paper of the ranking is labelled with a group of pupils or a viewpoint giving out superior than the programme.

IV. EXISTING SYSTEM

The current framework utilizes the decision tree system . These are non-parametric learning method utilized for characterization. And it is the one of most powerful techniques in classification .It constructs the models in the form of trees structure .Here datasets breaks into small sets at the same time ,an associated decision tree is formed. It can hold both the numeric data and cat goral. It is taken based on the attributes.

Drawbacks of Existing framework

1. High Cost.
2. Difficult to handle for non numeric data.
3. Many branches is difficult to understand with trees.
4. Tedious .
5. It is less expert
6. Time consuming.

V. PROPOSED SYSTEM

In this Proposed work is to construct model for forecasting of the set base on certain attributes. The work applies the following algorithms: Naive Bayes algorithm to classify and develop a model to make a diagnosis heart attacks in the patient data set from medical practiti0ners .Naive Bayes is the Probabilistic classifier based on the independent feature model. In general it assumes that presence or absence of any other characteristic trained to work in supervised learning manner

Favorable circumstances

1. Effective of cost
2. From the given earlier attributes heart disease can be predicted.
3. elevated precision in expectation.
4. Helps wellbeing segment in anticipating the sickness prior.
5. Easily handle large amount of data.
6. Handles real and discrete data.

VI. RESULTS AND DISCUSSION

System Perspective

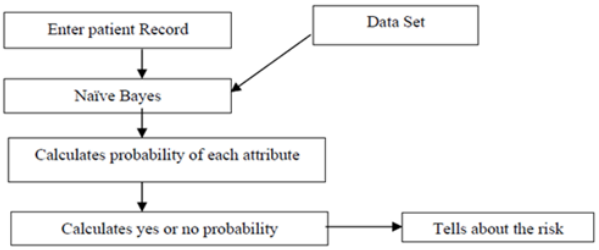


Figure:3 Architecture of Prediction of Heart disease

Detailed Design

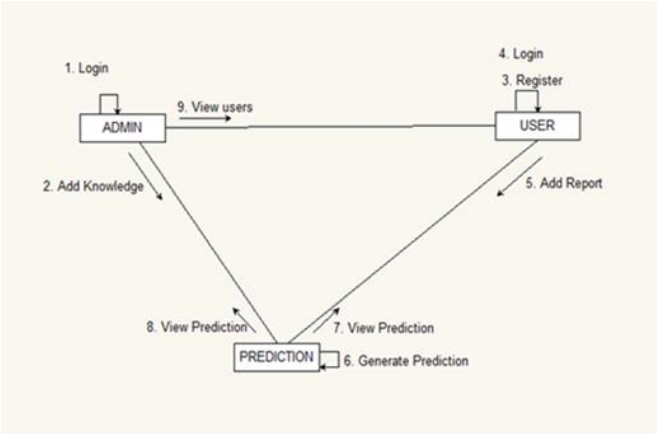


Figure 4. Celebration Diagram

Figure 5. Admin Login

Figure 6. User Registration

Figure 7. Admin has previous records

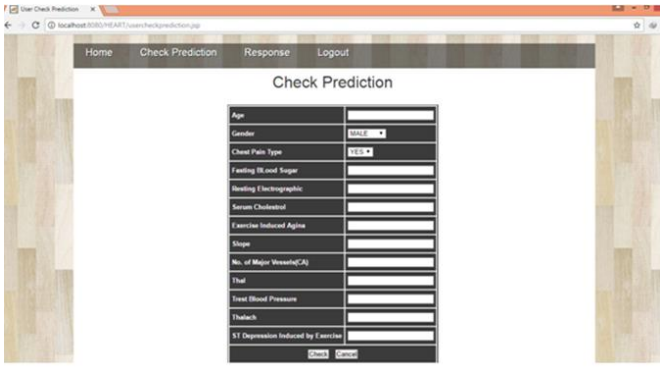


Figure 8. User Check Prediction

Figure 9. Doctor has parent details

VII. CONCLUSION

Data mining techniques are used to find out the hidden data for making it useful. Henceforth, the goal is met by utilizing the patients informational collection or qualities we can easily predict the possible heart attack from the given data set of the patients ..Many session of trials were led with the same datasets in UCI Machine Repository i,e.WEKA tool. Data set of some set of records with limited number of attributes is used and the outcome reveal that ,the Naïve Bayes beats the Decision Tree. Prognosis of the heart disease will be evaluated according to the result produced from it. Improvement is done to increase its consistency and efficiency.

VIII. FUTURE ENHANCEMENTS

Genetic algorithm will be used f future with the purpose of decrease the definite size of data to obtain most favourable division of point adequate for heart disease forecast. Prognosis of heart disease will be estimated according to the outcome produced from it. enhancement is done to increase efficiency and consistency. profit of genetic algorithm is the diagnosis of heart disease can be done in small time with the help of concentrated data sets.

IX. REFERENCES

[1]. Divya Kundra, Er. Navpreet Kaur, " Review 0n prediction system for heart diagnosis using Data Mining Techniques, IJLRET ISSN: 2454- 50, ISSN: 2454-5031

[2]. Ms.Rupali R.Patil, "Heart Disease Prediction System using Naïve Bayes and Jelinek-mercer smoothing", International Journal of Engineering Science and C0mputing, March 2016 2680 http://ijesc.org/ International J0urnal Advanced Research in Computer and Communication Engineering Vol. 3, Issue 5, May 2014.

[3]. K.Manimekalai, "A Review on Prediction of Heart Diseases by C0mparing Risk Factors in Data Mining", IJCSIT, ISSN:0975- 9646, Vol 7(1), 2016, 396-398.

[4]. B.Venkatalakshmi, M.V.Shivashankar, "Heart Disease Diagn0sis Using Predictive Data Mining", IJIRSET, Vol 3, March 2014, ISSN : 2319-8753

[5]. Hlaudi Daniel Masethe, Mosima Anna Masethe, "Prediction of Heart Disease using Classification Alg0rithms", Proceedings of the W0rld Congress on Engineering and Computer Science 2014 Vol II WCECS 2014, 22-24 October, 2014, San Francisco, USA.

[6]. Nidhi Bhatla and Kiran Jyoti, "An Analysis of Heart Disiease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278- 0181,Vol. 1 Issue 8, October – 2012

[7]. Abhishek Taneja, "Heart Disease Prediction System Using Data Mining Techniques", Oriental Journal of computer Science & Technology , An International Open Free Access, Peer Reviewed Research Journal.

[8]. Abhishek taneja, Heart Diseiase Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co., India, 2013.