Heart Diseases Prediction using Machine Learning

Mr.K.Venkata Ratnam[1] , P.Supriya[2]
[1]Assistant Professor, Dept of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.
[2]PG Scholar, Dept of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.

Abstract

The provision of healthcare is a necessity in human existence. The discipline of health concern business has grown to be significant in the expansive realm of medical science. The health care sector has a lot of data and secret information. By using data mining tools, judgments may be made with this concealed knowledge. In order to detect cardiovascular problems in a patient, numerous tests are performed; however, data mining could reduce the number of these tests. In order to classify the data and identify cardiac illnesses, a system is designed utilising data mining techniques because there is a shortage of analysis tools that can deliver accurate test results with the concealed information. The use of data mining can help with a variety of healthcare issues. One such data mining method that aids in the diagnosis of patients with cardiac problems is the naive bayes algorithm. This study examines a few variables and makes predictions about heart disease; it then proposes an HDPS based on data mining techniques.

Keywords: *healthcare system, datamining, heart diseases, naïve bayes algorithm, heart diseases prediction system.*

1. Introduction

The goal of the health care system is the preservation of one's bodily and emotional well-being by aiding in illness prevention or treatment. Heart disease is the leading global killer in the modern world. According to the World Health Organization, cardiac disorders are responsible for 12 million deaths worldwide each year. Heart disorders are responsible for more than 80% of deaths worldwide. Who predicted that by 2050, about 23.6 million people would die from heart disease. Adults with moderate to severe native heart disorders, infectious endocarditis, and prior valve intervention were included in the 25-nation Euro Heart Survey on Heart Diseases. 71.9 percent of patients had heart disease naturally, and 28.1 percent had previously undergone an intervention. The range of ages was 64 to 14 years. The majority of mitral stenosis causes in this country are rheumatic in origin, while degenerative aetiologies are the most common in aortic heart disorders and mitral regurgitation. Intelligent medical health care systems have benefited greatly from data mining.

Medical data mining in healthcare is regarded as a crucial yet challenging task that must be carried out precisely and effectively. In order to diagnose and treat diseases, health care data mining aims to address real-world health issues.The heart disease prediction system, a computerised approach for detecting heart diseases based on prior data and information, can be used to examine the relationship between problems and the real cause of disorders as well as the consequences of symptoms that are exhibited spontaneously in patients.

2. RELATED WORK

Health care professionals view medical data mining as a crucial but challenging work that must be carried out precisely and effectively. Health care data mining aims to address current issues in disease detection and treatment. The heart disease prediction system, a computerised approach for detecting heart diseases based on prior data and information, can be used to examine the relationship between problems and the real cause of disorders as well as the consequences of symptoms that are exhibited spontaneously in patients.

FigTable 1: Table shows different data mining techniques used in the diagnosis of Heart disease over different Heart disease datasets

Author	Year	Technique used	Attributes
Dr.K.Usharani	2011	Clasification/Neural Networks	13
Jesminahar,et al	2013	Apriori/Predictive Apriori/Tertius	14
Latha,et al	2008	Genetic Algorithm/CANFIS	14
Majabber,et al	2011	Clustring/Association Rule	14

		Mining/Sequence number	
Ms.Lshtake et al.	2013	DecisionTree/Neural Network/Naive bayes	15
Nan-Chen et al.	2012	(EVAR)/Machine Learning/Markov blanket	
Oleg et al.	2012	ANN/Genetic Poliymorphisms	
Shadab et al	2012	Naive bayes	15
Shantakumar et al.	2009	MAFIA/Clustring/K-Means	13
Carlos et al	2001	Association Rule	25

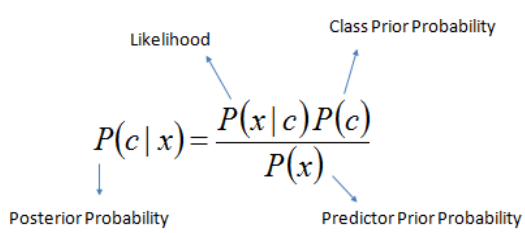
3. METHODOLOGY

3.1 Naive Bayes:

Naive Bayes classifiers is a probabilistic classifiers based onapplying B ayes' theorem w ith strong(naive) independence a ssumptions between the features. A Naive Bayesian model is simple to construct and does not require time-consuming iterative parameter estimation, making it very beneficial in the field of medicine for identifying heart patients. The Naive Bayesian classifier is popular because, despite its simplicity, it frequently outperforms more complex classification techniques.

From P(c), P(x), and P(x|c), the posterior probability, P(c|x), can be calculated using the Bayes theorem. The Naive Bayes classifier makes the assumption that the impact of a predictor's value (x) on a particular class (c) is unrelated to the values of other predictors. It is known as class conditional independence.

3.2 Equations:



$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- $P(c|x)$ is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
- $P(c)$ is the prior probability of *class*.
  of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*

Where C and X are two events (e.g. the probability that the train will arrive on time given that the weather is rainy). Such Naïve Bayes classifiers use the probability theory to find the most likely classification of an unseen (unclassified) instance . The algorithm performs positively with categorical data but poorly if we have numerical data in the training set.

4 .Dataset

The clinical data set used in this study was gathered from one of the top centres for diabetes research in Chennai and includes information on roughly 500 patients. For items linked to diabetes, the clinical data set specification provides clear, succinct definitions. The goal of the diabetes data set is to provide people with diabetes with up-to-date records of their risk factors, current management, treatment target accomplishments, and plans for and results of routine surveillance for complications. This will allow them to monitor their care and make wise management decisions. Additionally, it will guarantee that when persons with diabetes consult with medical specialists, the consultation is fully supported by thorough, current, and accurate information. The diabetes attributes used in our proposed system and their descriptions are shown in FigTable 2.

FigTable 2: Parameters of Heart Diseases Prediction System

Name	Type	Description
Age	Continuous	Age in years
Sex	Discrete	1 = male 0 = female
Cp	Discrete	Chest pain type: 1 = typical angina 2 = atypical angina 3 = non-angina pa 4 =asymptomatic
Trestbps	Continuous	Resting blood pressure (in mm Hg)
Chol	Discrete	Serum cholesterol in mg/dl
Fbs	Discrete	Fasting blood sugar > 120 mg/dl: 1 = true

		0 = false
Restecg	Continuous	Resting electrocardiographic results: 0 = normal 1 = having ST-T wave abnormality 2 =showing probable or define left ventricular hypertrophy by Estes 'criteria
Thalach	Discrete	Maximum heart rate achieved
Exang	Discrete	Exercise induced angina: 1 = yes 0 = no
Slope	Discrete	The slope of the peak exercise segment : 1 = up sloping 2 = flat 3= down sloping
Diagnosis	Discrete	Diagnosis classes: 0 = healthy 1= possible heart disease

5 .Data mining Tool

For data mining jobs, Weka is a collection of machine learning algorithms. You can either invoke the algorithms directly from your own Java code or apply them directly to a dataset. Data pre-processing, classification, regression, clustering, association rules, and visualisation tools are all included in Weka. It is also suitable for creating novel machine learning techniques. The weka tool is used to conduct the experiments, and the outcomes are obtained. To do classification utilising a 70% percentage split, we employed the Navier-Bayes algorithm.

6.Data Analysis:

In this system the medical data set is classified based on the classes present/absent.The proposed naïve bayes model was able to classify 86.4198% of the input instances correctly and the incorrect instances was 13.5802% for 70% of percentage split. With the total of 81 instances 70% was classified as correct and 11% instances was incorrect.The results clearly states that naive bayes provides better results regarding the people affected by heart diseases.

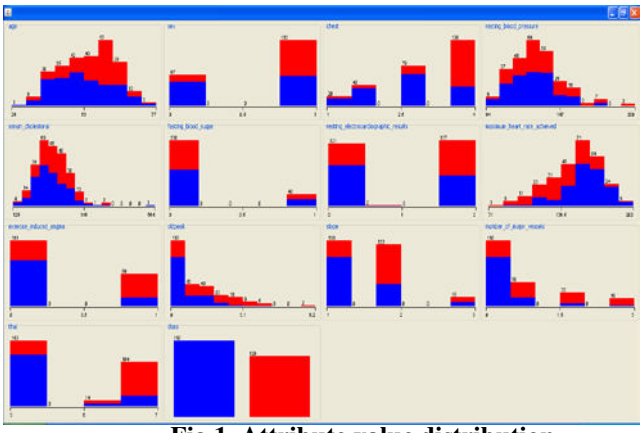


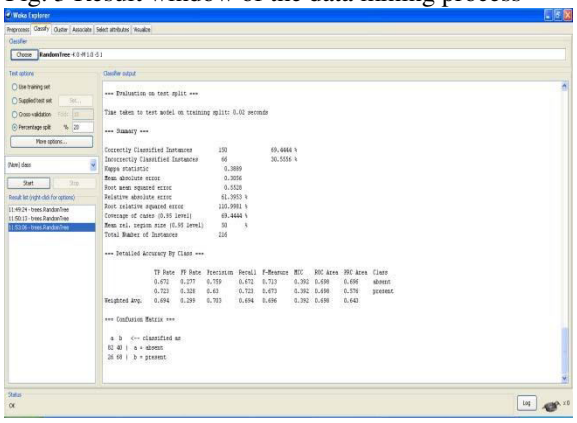
Fig.1. Attribute value distribution

The blue colored regions in the graphs in Figure 1 denote high cholesterol values. The graphs show that the majority of diabetic patients with high cholesterol values are between the ages of 45 and 55, have a body weight between 60 and 71 pounds, a blood pressure reading of 148 or 230, a fasting blood sugar reading between 102 and 135 pounds, a PP reading between 88 and 107 pounds, and an A1C reading between 7.7 and 9.6 pounds.

7 .Results and Discussions:

The results of our experimentation are shown in Fig2.

Fig. 3 Result window of the data mining process



The suggested naive bayes model properly classified 74% of the input cases. It displayed an average F-measure of 71.2 percent, accuracy of 71%, and recall of 74%. The results clearly demonstrate that the suggested technique outperforms other comparable methods in the literature, even when it is taken into account that the features used for analysis are not specific indications of heart disease..

8.Conclusion:

Applications for data mining are widely utilised in the medical profession to identify ailments and diagnose cardiac patients based on the given data set and qualities. The use of various data mining techniques to aid medical practitioners in the detection of heart disease has been studied by researchers. Because the Navie Bayes method produces precise findings, it is employed in the proposed study to categorise the data set and estimate the prevalence of heart disease in humans. Heart illnesses are therefore accurately predicted by the heart diseases prediction system using the medical data. The outcomes demonstrate that the Navie Bayes algorithm offers 86.4198 percent accuracy in the shortest amount of time.

References

[1] S. K. Yadav, B. K. Bharadwaj & Pal, S. 2011. Data Mining Applications: A comparative study for predicting students' performance, International journal of Innovative Technology and Creative Engineering (IJITCE), 1(12).

[2] Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.

[3] Wu, X., et al., Top 10 algorithms in data mining analysis. Knowl. Inf. Syst., 2007.

[4] S. K. Yadev & Pal., S. 2012. Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification, World of Computer Science and Information Technology (WCSIT), 2(2), 51-56.

[5] Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.

[6] Srinivas, K., B.K. Rani, and A. Govrdhan, Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. International Journal on Computer Science and Engineering (IJCSE), 2010. Vol. 02, No. 02: p. 250-255.