IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, September 2015.
www.ijiset.com
ISSN 2348 – 7968

Heart Diseases Detection Using Naive Bayes Algorithm

K.Vembandasamy[1], R.Sasipriya[2]and E.Deepa[3]

[1] Department of computer science, PSG college of arts and science, Coimbatore, Tamilnadu, India,

[2] Department of computer science, PSG college of arts and science, Coimbatore, Tamilnadu, India,

[3] Department of computer science, PSG college of arts and science, Coimbatore, Tamilnadu, India,

Abstract

Health care is an inevitable task to be done in human life. Health concern business has become a notable field in the wide spread area of medical science. Health care industry contains large amount of data and hidden information. Effective decisions are made with this hidden information by applying dataminnig techniques. Several tests are done in the detection of cardiovascular diseases in the patient; however with datamining these tests could be reduced. But there is a lack of analysing tool to provide effective test results with the hidden information, so a system is developed using data mining algorithms for classifying the data and to detect the heart diseases. Datamining acts as a solution for many healthcare problems. Naïve bayes algorithm is one such datamining technique which serves in the diagnosis of heart diseases patient. This paper analyse few parameters and predicts heart diseases, there by suggests a heart diseases prediction system (HDPS) based on the datamining approaches.

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

1. Introduction

Health care system is the preservation of mental and physical heath by preventing or treating illness through services offered by the provision. In day today life heart diseases is the major cause of deaths in the world. The world health organization (who) has estimated that 12 million death occurred world wide, each year due to heart diseases. Over 80% of deaths in world are because of heart diseases. Who estimated in future, almost 23.6 million people will die due to heart diseases. The euro heart survey on heart diseases was conducted by 25 countries it included adults with moderate severe native heart diseases, infective endocarditic are previous valve intervention. Heart diseases was native in 71.9% of patients and 28.1% had had a previous intervention. Mean age was 64 to 14 years. Degenerative aetiologies where the most frequent in aortic heart diseases and mitral regurgitation while most cause of mitral stenosis

Here of rheumatic origin. Data mining has been played an important role in the intelligent medical health care systems. Medical data mining in health care is regarded as an important yet complicated task that needs to be executed accurately and efficiently. Health care data mining attempts to solve real world health problems in diagnosis and treatment of disease. The relationship of disorders and real cause of disorders and the effects of symptoms that are spontaneously seen in patients can be evaluated by using the heart disease prediction system, is a computerized method for diagnosing heart diseases based on prior data and information.

2. RELATED WORK

Heart disease is a term that assigns to a large number of medical conditions related to heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its parts. Heart disease is a major health problem in today's time. This paper aims at analyzing the various data mining techniques introduced in recent years for heart disease prediction. Table 1 shows different data mining techniques used in the diagnosis of Heart disease over different Heart disease datasets. In some papers this is given that they use only one technique for diagnosis of heart disease as given in Shadab et al , Carlos et al etc. but in case of other research work more than one data mining techniques are used for the diagnosis of heart disease as given in Ms. Ishtake et al., MA.JABBAR, et al, Shantakumar et al etc.

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

ISSN 2348 – 7968

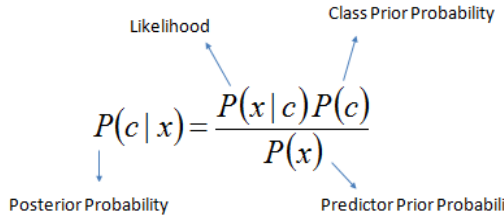
		Mining/Sequence number	
Ms.Lshtake et al.	2013	DecisionTree/Neural Network/Naive bayes	15
Nan-Chen et al.	2012	(EVAR)/Machine Learning/Markov blanket	
Oleg et al.	2012	ANN/Genetic Poliymorphisms	
Shadab et al	2012	Naive bayes	15
Shantakumar et al.	2009	MAFIA/Clustring/K-Means	13
Carlos et al	2001	Association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

3. METHODOLOGY

3.1 Naive Bayes:

Naive Bayes classifiers is a probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful in the field of medical science for diagnosing heart patients. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.
Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence

3.2 Equations:



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

- $P(c|x)$ is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability 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 data set used in this work are clinical data set collected from one of the leading diabetic research institute in Chennai and contain records of about 500 patients. The clinical data set specification provides concise, unambiguous definition for items related to diabetes. The diabetes data set is developed to ensure people with diabetes have up to date records of their risk factors, current management, treatment target achievements and arrangements and outcomes of regular surveillance for complications, to help them monitor their care and make informed choices about their management. It will also ensure that when people with diabetes meet health care professionals the consultation is fully informed by comprehensive, up to date 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

ISSN 2348 – 7968

		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

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. The experiments are conducted using the weka tool and the results are obtained. We have used the navie bayes method to perform classification by using 70% of percentage split.

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.

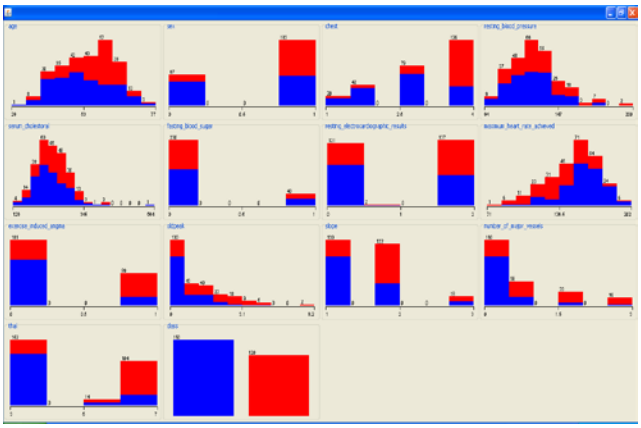


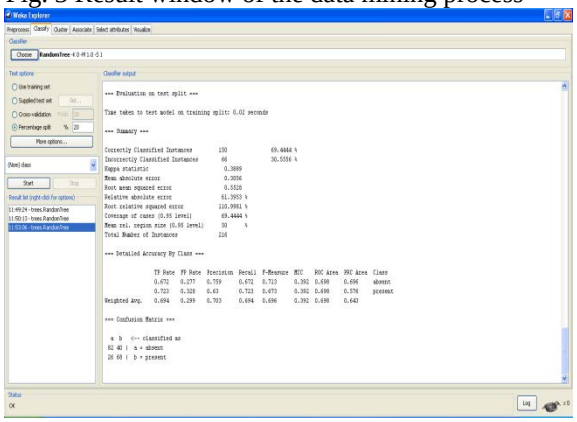
Fig.1. Attribute value distribution

The blue colored regions in the graphs in Figure 1 denote high cholesterol values. From the graphs we can see that, most of the diabetic patients with high cholesterol values are in the age group of 45 – 55, have a body weight in the range of 60 – 71, have BP value of 148 or 230, have a Fasting value in the range of 102 – 135, have a PP value in the range of 88 – 107, and have a A1C value in the range of 7.7 – 9.6.

7 .Results and Discussions:

The results of our experimentation are shown in Fig2.

Fig. 3 Result window of the data mining process



The proposed naïve bayes model was able to classify 74% of the input instances correctly. It exhibited a precision of 71% in average, recall of 74% in average, and F-measure of 71.2% in average. The results show clearly that the proposed method performs well compared to other similar methods in the literature, taking into the fact that the attributes taken for analysis are not direct indicators of heart disease.

ISSN 2348 – 7968

8 .Conclusion:

Data mining applications are used vastly in the medical field to detect diseases and diagnosis the heart patient based on the data set and the attributes provided. Researchers have been investigating applying different data mining techniques to help health care professionals in the diagnosis of heart disease. In the proposed work navie bayes algorithm is used to classify the data set because navie bayes provides accurate results, with these results heart diseases among people is predicted. Thus heart diseases prediction system successfully diagnose the medical data and predicts the heart diseases.The results thus obtained shows that navie bayes algorithm provides 86.4198% of accuracy with minimum 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.