© 2020 IJSRST | Volume 5 | Issue 7 | Print ISSN : 2395-6011 | Online ISSN: 2395-602X

4th National Conference on Advances in Engineering and Applied Science
Organized by : Anjuman College of Engineering and Technology (ACET) Nagpur, Maharashtra, India, In association with
International Journal of Scientific Research in Science and Technology

Heart Disease Prediction using Machine Learning Algorithms

Khan Afreen Wahab*, Prof. Dr. A. S. Khan
Electronics and Telecommunication, RTM University, Nagpur, Maharashtra, India

ABSTRACT

Data Mining is a technique to recognize proposals of information or choice making awareness in the database and extracting these to be used in different areas such as decision support systems, predictions, and forecasting and estimation. The medical field collects huge amounts of Medical data which regrettably, are not "mined" to find out concealed information for successful decision making. This study has urbanized a verdict Support in Heart Disease Prediction System (DSHDPS) by means of data mining and modeling method, namely, Naïve Bayes. By considering 15 special medical profiles of 303 patients and achieved accuracy of 82.5 % which is better than some other classifiers. Naive Bayes can envisage the likelihood of patients getting a heart disease.

Keywords: Data Mining, Decision Support, Heart Disease, Naïve Bayes.

I. INTRODUCTION

Data Mining technique used in finding valid, novel, and potentially useful Technique and finally easily known pattern in data with the more use of databases and the Data mining is the search for the links and global patterns that remain in huge databases but are masked among bigger amounts of data. The important process of wisdom Discovery is the alteration of data into wisdom in order to aid in conclusion, referred to as data mining.

In Many hospitals information systems are designed in keeping view of patient billing, management, generation of simple statistics things. "When a patient data is given, predict the chances of patients obtaining a heart disease." Clinical related decisions are made based on doctors' incitation and experience moderately than on the ability rich data masked in the database. This method bulge to unnecessary biases, faults and more medical costs which involve the aspects of service given to patients. The designed system that incorporate of medical decision which is supported by means of computer-based patient records may break medical faults, improve patient safety, reduce unnecessary practice variation, and enhance patient outcome. This approach is assuring as data designing and testing tools, e.g., ML , have the ability to produce a knowledge rich medium which can help to consequently enhance the quality of Medical decisions.

Now-a –days it is a must for all the hospitals with more flow of patients to keep track of patient data there is a wealth of concealed information in these data which is largely unused. Data is curved into practical information that is useful for doctors and healthcare practitioners to formulate intelligent medical decisions. The main purpose of this intended study is to enlarge a Decision Support in Heart Disease Prediction System (DSHDPS) using data mining technique, called as Naïve Bayes We are able to give the report of the patient in 2 ways by using1) chart and the second one as 2) pdf which will tell u whether the intended person has the heart disease or not. Data mining, have the prospective to make a information rice environment which is used

to considerably get better excellence of clinical decisions.

II. METHODS AND MATERIAL

Clinical databases have collected large amount of information about patients and their medical situation. The term Heart disease encompasses the various diseases that influence the heart. Heart disease is the major reason of sufferers in the world. Due to Heart disease one person expires for every 34 seconds in U S. The three different forms of heart disease are 1) Coronary heart disease,2) Cardiomyopathy and 3) Cardiovascular disease. The term "cardiovascular disease" include a extensive range of circumstances that concern the heart and the blood vessels and the way in which blood is pumped and circulated through the body. Cardiovascular disease (CVD) results in cruel illness, disability, and leads to death. The data for our research is obtained from the Cleveland Heart Disease database which has 15 attributes and with 303 patients . The records were split into two datasets: 70 percent as training dataset and 30 percent as testing dataset. The attribute "Diagnosis" is recognized as the expected attribute with value "1" for patients with heart disease and value "0" for patients with no heart disease.

Key attribute

Input attributes

1. Age in Year
2. Sex (value 1: Male; value 0: Female)
3. Chest Pain Type (value 1:typical type 1 angina, value 2: typical type angina, value 3:non-angina pain; value 4: asymptomatic)
4. Fasting Blood Sugar (value 1: >120 mg/dl; value 0: <120 mg/dl)
5. Restecg – resting electrographic results (value 0:normal; value 1: having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy)
6. Exang - exercise induced angina (value 1: yes; value 0: no)
7. Slope – the slope of the peak exercise ST segment (value 1:unsloping; value 2: flat; value 3: downsloping)
8. CA – number of major vessels colored by floursopy (value 0-3)
9. Thal (value 3: normal; value 6: fixed defect; value 7: reversible defect)
10. Trest Blood Pressure (mm Hg on admission to the hospital)
11. Serum Cholestrol (mg/dl)
12. Thalach – maximum heart rate achieved
13. Oldpeak – ST depression induced by exercise
14. Smoking – (value 1: past; value 2: current; value 3: never)
15. Obesity – (value 1: yes; value 0: no)

III. RESULTS AND DISCUSSION

Confusion Matrix:

A. A confusion matrix is used to indicate the performance of a ML classifier on a set of given test data . It tells about the performance of an algorithm.

TABLE 1. Confusion Matrix

TP	FN
FP	TN

1) True Positive (TP) : Observation is positive, and is predicted to be positive.
2) False Negative (FN) : Observation is positive, but is predicted negative.
3) True Negative (TN) : Observation is negative, and is predicted to be negative.
4) False Positive (FP) : Observation is negative, but is predicted positive.

Classification Rate/Accuracy:

Classification Rate or Accuracy is given by the relation:

Predicative Accuracy : TP+TN / (TP+FN+FP+TN)

Sensitivity = Number of true positives
Total number of individuals with the illness

The specificity of a test is the probability that a test gives a negative result given that the patient will not have the disease.

Specificity = Number of true negatives/Total number of individuals without the illness

TABLE 2. Comparative Analysis

Algorithm	**Predicative Accuracy**	**Correctly Classified Instances**	**Sensitvity**	**Specificity**
Naïve Bayes	**82.5**	**250**	**0.85**	**0.79**
Tree	72.6	220	0.72	0.73
SVM	82.5	250	0.89	0.73
Logistic Regression	81.5	247	0.85	0.76

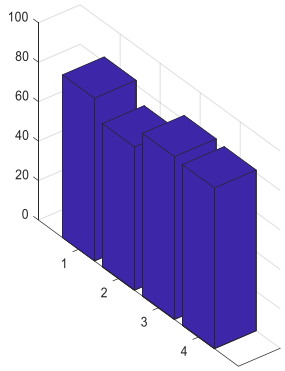

Figure 1 : Predicative Accuracy Vs Algorithms (Naivye Bayes, Desicion Tree, SVM, Logistic Regression)

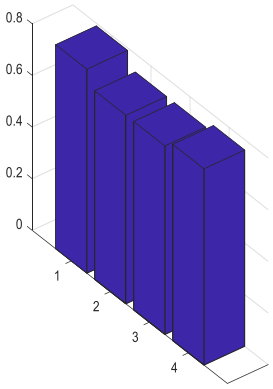

Figure 2 : Sensitivity Vs Algorithms (Naivye Bayes, Desicion Tree, SVM, Logistic Regression)

CONCLUSION

DSHDPS is elaborated by using Naive Bayesian ML technique. The system extracts knowledge from heart disease database. This is the most successful model to estimate patients with heart disease. This model could give solution to difficult queries, refere to every one wwith its own quality w.r.t to model explanation, access to complete data and accuracy. It can be further intensify and elaborated. It can take in to consideration other medical attributes also. It can also use other ML techniques. We can use all forms of Data like1) numeric 2)categorical and 3)both.

IV. REFERENCES

[1]. Blake, C.L., Mertz, C.J.: "UCI Machine LearningDatabases",, 2004.

[2]. Chapman, P., Clinton, J., Kerber, R. Khabeza, T., Reinartz, T., Shearer, C., Wirth, R.: "CRISP-DM

[3]. Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", 978-1-4244-1968- 5/08/$25.00 ©2008 IEEE.

[4]. Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006

[5]. Ho, T. J.: "Data Mining and Data Warehousing", Prentice Hall, 2005.

[6]. Sellappan, P., Chua, S.L.: "Model-based Healthcare Decision Support System", Proc. Of Int. Conf. on Information Technology in Asia CITA'05, 45-50, Kuching, Sarawak, Malaysia, 2005

[7]. Tang, Z. H., MacLennan, J.: "Data Mining with SQL Server 2005", Indianapolis: Wiley, 2005.

[8]. Shantakumar B.Patil, Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research ISSN 1450-216X Vol.31 No.4 (2009)

[9]. Wu, R., Peters, W., Morgan, M.W.: "The Next Generation Clinical Decision Support: Linking Evidence to Best Practice", Journal Healthcare Information Management. 16(4), 50-55, 2002.

[10]. Kaur, H., Wasan, S. K.: "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 194-200, 2006