International Journal of Advance Research in Science and Engineering
Vol. No.9, Issue No. 03, March 2020
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Implementation of classification algorithms on heart disease dataset

[1]P.Nithya M.Sc., M.Phil., [2]C.Rajeswari ,[3]D.Shanmugapriya M.Sc(IT)

Department of Computer Science Information Technology,
Nadar Saraswathi College of Arts & Science, Theni, India

Abstract

Data mining Techniques has proved for early prediction of disease with higher accuracy in order to save human life. Heart disease diagnosis is a complex task which requires much experience and knowledge. It is a single largest cause of death in developed countries. In this paper two classification algorithms, namely Naive Bayesian, J48 studied and applied on the heart disease dataset. The so- called algorithms perform diagnosis using WEKA tool and compare its accuracy rate.

Keywords: Data mining, Heart disease Dataset, Naive Bayesian, J48.

1. INTRODUCTION

1.1 Data mining

Data mining has already established as a novel field for exploring hidden patterns in the huge datasets. Medical science is another field where large amount of data is generated using different clinical reports and other patient symptoms. Data mining can also be used heavily for the same purpose in medical datasets also. These explored hidden patterns in medical datasets can be used for clinical diagnosis. However, medical datasets are widely dispersed, heterogeneous, and huge in nature.

1.2. Heart Disease

Heart attack diseases remains the main cause of death worldwide, an earlier stage will prevent the attacks. Medical practitioners generate data with a wealth of hidden information present, and it's not properly being used effectively for predictions [2]. For this purpose, the research converts the unused data into a dataset for modeling using different data mining techniques. People die having experienced symptoms that were not taken into considerations. There is a need for medical practitioners to predict heart disease before they occur in their patients. The features that increase the possibility of heart attacks are smoking, lack of physical exercises, high blood pressure, high cholesterol, unhealthy diet, harmful use of alcohol, and high sugar levels[3].

There are number of factors which increase the risk of Heart disease:

- Family history of heart disease
- Smoking
- Cholesterol
- Poor diet
- High blood pressure

- High blood cholesterol
- Obesity
- Physical inactivity
- Hyper tension

Symptoms of a Heart Attack

Symptoms of a heart attack can include:

- Discomfort, pressure, heaviness, or pain in the chest, arm, or below the breastbone.
- Discomfort radiating to the back, jaw, throat, or arm.
- Fullness, indigestion, or choking feeling (may feel like heartburn) [4].

2. Related Work

The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al. [5]. The dataset contains records with 13 attributes in each record. The supervised networks i.e. Neural Network with back propagation algorithm is used for training and testing of data.

The researchers proposed a layered neuro-fuzzy approach to predict occurrences of coronary heart disease simulated in MATLAB tool. The implementation of the neuro-fuzzy integrated approach produced error rate very low and high work efficiency in performing analysis for coronary heart disease occurrences [6].

Another study experimented on a sample database of patients' records. The Neural Network is tested and trained with 13 input variables such as Age, Blood Pressure, Angiography's report and the like. The supervised network has been recommended for diagnosis of heart diseases [5]

3. Methodology

3.1Naive Bayesian

This algorithm provides a prediction model in relation to the likelihood of certain outcomes. Naive Bayesian algorithm measures patterns or relationships among data by counting the number of observations. The algorithm then creates a model that reflects the patterns and their relationships. After creating this model, it can be used as a prediction of several objectives.

3.2 J48

It is a decision tree that uses the concept of entropy with a training dataset. The decision tree is a method to display a series of rules, leading to a class or value. In J48 algorithm, every feature of the data is used to make a decision by splitting into smaller subsets. J48 uses a statistical value called the Information Gain to determine how much a property can separate the training data according to their classification. The information interest of a feature is the amount of entropy reduction that can be achieved by separating data through this feature.

4. Results and Discussion

Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D. -- 2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D. -- 3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D. -- 4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D. (b) Donor:

David W. Aha (aha@ics.uci.edu) (714)856-8779 (c) Date: July, 1988. It has 14 attributes, 304 instances .In that we have taken 7 clinical attributes and 30 sample Instances.The are age,sex,cp,oldpeak,slope,thal,num.
The correctly classified instances are shown in Table1.

Table 1. Classified Instances

Dataset	Algorithm	Correctly classified instance
Heart dataset 130 Hungarian Hospitals…	Naïve Bayesian	62%
	J48	55%

5. Conclusion

The automatic diagnosis of Heart Disease is an important real-world medical problem. Detection of Heart Disease in its early stages is the key for treatment. In this work, we have compared two classification algorithm results. Finally Naïve Bayes produced 62% of correctly classified instances when compared to the J48 algorithm. In future study the work can be extended and improves for the automation of Heart Disease.

6. Reference

[1] David L. Olson and Dursun, D., *Advanced Data Mining Techniques*. Springer-Verlag Berlin Heidelberg (2008).

[2] S. Ishtake and S. Sanap, "' Intelligent Heart Disease Prediction System Using Data Mining Techniques ',"" *International Journal of healthcare & biomedical Research*, vol. 1, no. 3, pp. 94–101, 2013.

[3] D. S. Chaitrali and A. S. Sulabha, "A Data Mining Approach for Prediction of Heart Disease Using Neural Networks," *International Journal of Computer Engineering & Technology (IJCET)*, vol. 3, no. 3, pp. 30–40, 2012.

[4]K.Sudhakar, Dr. M. Manimekalai " Study of Heart Disease Prediction using Data Mining "*International Journal of Advanced Research in Computer Science and Software EngineeringVol 4,Issue 1,2014*

[5] Niti Guru, Anil Dahiya, NavinRajpal, *"Decision Support System for Heart Disease Diagnosis Using Neural Network"*, Delhi Business Review, Vol. 8, No. 1 (January - June 2007).

[6] A. K. Sen, S. B. Patel, and D. P. Shukla, "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level," *International Journal of Engineering and Computer Science*, vol. 2, no. 9, pp. 1663–1671, 2013.