International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 11 6090 - 6092

Detection of Coronary Heart Diseases using Data Mining Techniques

Harita Jagad, Jehan Kandawalla, Prof. Sindhu Nair
Department of Computer Engineering
Dwarkadas Jivanlal Sanghvi College of Engineering

Abstract- In recent times, heart diseases have been prevalent throughout the globe. Millions of people die every year because of misdiagnosis of heart diseases. There is a dire need to develop correct diagnosis of heart related diseases based on data from earlier case scenarios.To aid early and correct diagnosis of these Coronary Heart Diseases, many data mining techniques have been devised using large amount of data available to the hospitals. The various algorithms devised are used to assist physicians in the diagnosis of heart diseases based on various patient parameters such as fasting blood sugar, previous heart events, age, etc. This paper analyses the best of these techniques which include the decision tree algorithm, the Naïve Bayes algorithm and the neural networks algorithm, and enlists the good and the bad in each one of them.

Keywords--*Coronary Heart Disease; Decision trees; Naïve Bayes; Neural Networks*

*****

I. INTRODUCTION:

In this paper we are using three data mining algorithms, decision tree, Naïve Bayes and neural network which will be used to classify data sets based on patient history and assist physicians in the diagnosis of coronary heart diseases. Thus mining patient data and eventually using this for prediction of possible heart disease in the forthcoming data can save a number of lives.

- In *decision trees* historical data is used to form a conditional tree as a machine learning method. Then the algorithm will start from the root node each time and proceed step by step based on the set of patient parameters entered until it reaches any end node. This end node will show the probability of the person suffering from a given coronary heart disease.
- *Naïve Bayes* algorithm uses various patient parameters and then finds the probability of a the patient having a coronary heart disease based on every individual parameter.These individual probabilities may later on be combined by means of some formulae to find the probability of the patient suffering from a coronary heart disease based on all the parameters combined.
- *Neural networks* help us train large amount of patient data even if the input symptoms are less. Since Neural Networks give more importance to inputs with higher weights, which makes the more prominent parameters of a greater value.

II. ALGORITHMS:

In this section, we proceed to explain each of these three data mining techniques:

2.1Decision Tree:

Decision trees are developed by breaking down of large data sets into subset but at the same time they also look after the association between these data sets. They are traversed in a top-down approach.

In the Learning phase of this algorithm, a set of historical data is taken and using this data a conditional tree is formed which will satisfy all the constraints given in the data. After this the algorithm can be used to find the probability of a patient suffering from a given heart condition by means of a set of steps. The patient parameters like age, sex, fasting blood sugar level, thalach level, drinking and smoking habits, and a set of other symptoms(which are predefined in the decision tree) are taken during patient check-up. This data is entered into the algorithm. The algorithm then starts from the root node and proceeds to each corresponding node based on the value of a certain parameter. This process will continue till the patient parameter set is exhausted and a leaf node is reached. Along every step the algorithm gets closer to a diagnosis and the finaldiagnosis and its probability is shown by the value at the leaf node.

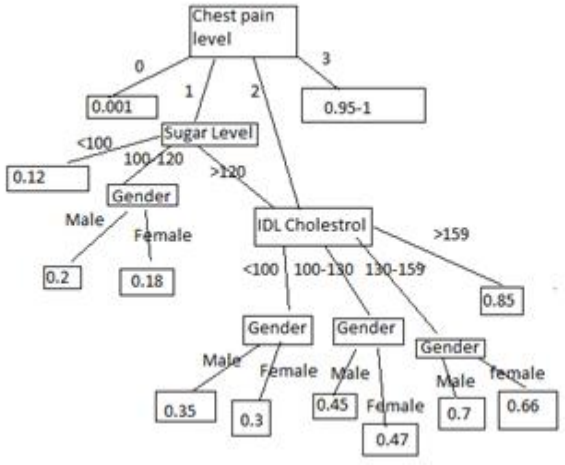

Figure 1- Decision tree

2.2 Naïve Bayes:

This algorithm determines a separate probability of the person suffering from a certain coronary heart disease based on a certain individual parameter. The patient parameters are entered into the system. After this the algorithm takes every single parameter and attributes a probability to it based on the information extracted from the historical data available on the system. Calculation of the probability of there being a heart disease based on any parameter is completely independent of the probability of the same for any other parameter. Once each individual different weightage depending on the importance of that parameter, and hence the final probability will be a cumulative function of the average of all values of weightage times probability.Once each individual probability is determined, we can devise a certain formula to derive the probability of there being a coronary heart disease based on all the parameters combined. Each parameter could be given a different weightage depending on the importance of that parameter, and hence the final probability will be a cumulative function of the average of all values of weightage times probability.

Cumulative probability=$\sum[(n-N)*P(W-N)]/\sum[n-N]$

N=0 to d

Where,W-0, W-1, W-2…..W-d are the different parameters and n-0, n-1….n-d is the weightage given to each of these parameters respectively.

2.3 Neural Network:

Neural Networks can use their hidden units to learn from derived features of the task, which is classification of patient data and use these
features learnt from the data set are used to
classify forthcoming inputs.This method is preferred because of its ability to learn from examples. Inputs to the Neural Network will be parameters like sex, cholesterol, age, etc. After training for numerous iterations, till the error is minimized, we obtain the actual classification. The fresh data that will be fed, will be easily classified with the help of the previously fed training data we got from the various hospitals. This process works in various iterations. Each input is initially assigned a certain random weightage. After this the output is computed. Then the output is checked with the real set of desired or ideal output. The difference between the two is known as error. The algorithm aims at removing or minimizing this error with each iteration by adjusting the weights of each parameter by means of mathematical formulae. Hence neural networks take time to compute but will arrive at the ideal conclusion even with very few input conditions.

III. DISCUSSION:

- The algorithms discussed in the study were decision trees, Naïve Bayes and Neural networks.Since the reliability of decision trees depends on the precision of input data from the very beginning, this might prove to be a little difficult while dealing with large sets of data. Though the data used will be from reliable sources and highly monitored, this might prove to be a disadvantage. On the other hand, decision tree has an advantage of having comprehensive data.
- Naïve Bayes needs to be combined with other statistical techniques to classify the best features, but it can easily run on large sets of data because all it needs is the learned probability of a heart even with respect to each individual parameter and then just uses one simple formulae to combine these conditional probabilities to find the overall probability of a coronary heart disease occurring in a person with a certain set of parameters. The Naive Bayes algorithm is usually the fastest of the three discussed in this paper, but it needs a moderate sized input set.
- Neural networks are basically used when the input sets are small. This method has a higher time complexity but it is the only method which would give close to ideal results with a limited input set. The scope for error is minimal as almost all the errors are removed by the numerous iterations.

Model Type	Prediction Attributes	No. of Cases	Prediction
Decision Tree	+WHD,+PHD	146	Correct
	-WHD,+PHD	27	Incorrect
	-WHD,-PHD	219	Correct
	+WHD,-PHD	62	Incorrect
Naïve Bayes	+WHD,+PHD	180	Correct
	-WHD,+PHD	35	Incorrect
	-WHD,-PHD	211	Correct
	+WHD,-PHD	28	Incorrect
Neural Network	+WHD,+PHD	178	Correct
	-WHD,+PHD	35	Incorrect
	-WHD,-PHD	211	Correct
	+WHD,-PHD	30	Incorrect

+WHD stands for- Patients with heart disease

–WHD stands for-Patients with no heart disease

+PHD stands for- Patients predicted as having heart disease

–PHDstands for- Patients predicted as having no heart disease

FIGURE 2 - Result Table [4]

REFERENCES

[1] Studies on application of Support Vector Machine in diagnose of coronary heart disease -Yan Zhang * Fugui Liu Zhigang Zhao□Dandan Li Xiaoyan Zhou Jingyuan Wang.

[2] Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees- Minas A. Karaolis, Member, IEEE, Joseph A.Moutiris, Demetra Hadjipanayi, and Constantinos S. Pattichis, Senior Member, IEEE.

[3] Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems Eman Abu Khousa, Piers Campbell- Faculty of Information Technology

[4] Intelligent Heart Disease Prediction System Using Data Mining TechniquesSellappanPalaniappan, RafiahAwangDepartment of Information TechnologyMalaysia University of Science and TechnologyBlock C, Kelana Square, Jalan SS7/26 Kelana Jaya, 47301 Petaling Jaya, Selangor, Malaysia

[5] AKAMAS: Mining Association Rules Using a New Algorithm for theAssessment of the Risk of Coronary Heart Events- M. Karaolis, Student Member, IEEE, l.A.Moutiris, FESC, L. Papaconstantinou,C.S. Pattichis, Senior Member, IEEE.

[6] A Study on Classification Techniques in Data Mining – G. Kesavaraj, Assistant Professor Department of Computer Applications,Vivekanandha College Of Arts And Sciences For Women, Dr.S.Sukumaran Associate professor,Department of Computer Science,Erode Arts and Science College,Erode.

[7] Using Data Mining Techniques in Heart Disease Diagnosis and Treatment- Mai Shouman, Tim Turner, Rob Stocker School of Engineering and Information Technology University of New South Wales at the Australian Defence Force Academy

[8] Patil, S.B. and Y.S. Kumaraswamy, Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction. International Journal of Computer Science andNetwork Security (IJCSNS), 2009. VOL.9 No.2

[9] Kim, J., et al., A Novel Data Mining Approach to the Identification of Effective Drugs or Combinations for Targeted Endpoints— Application to Chronic Heart Failure as a New Form of Evidence based Medicine. Cardiovascular Drugs and Therapy, Springer 2005. 18 p. 483–489.2012.

[10] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, "Decision trees: An overview and their use in medicine," *J. Med. Syst.*, vol. 26, no. 5, pp. 445–463, 2002.

[11] Tan, G. & Cbye H., "Data mining applications in healthcare," *Journal of Healthcare Information Management.* Vol. 19, No.2,2004.