Volume 5, Issue 2, February 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Heart Disease Diagnosis System Using Apriori Algorithm

T.Georgeena.S. Thomas, Siddhesh.S. Budhkar, Siddhesh.K. Cheulkar, Akshay.B.Choudhary, Rohan Singh
Department of INFT, Atharva College of Engineering,
Mumbai, Maharashtra, India

Abstract— Cardiovascular disease is a class of diseases that involve the heart, the blood vessels (arteries, capillaries, and veins) or sometimes both. Cardiovascular disease refers to any disease that affects the cardiovascular system, mainly cardiac disease, and vascular diseases of the brain and kidney, and peripheral arterial disease. The tests you'll need to diagnose your heart disease depend on what condition your doctor thinks you might have at that time. No matter what type of heart disease you have, your doctor will perform a physical exam and ask about your personal and family medical history before doing any tests. The main purpose of our work is to reduce the number of attributes used in heart disease diagnosis that will result into the reduce in the number of tests which are required to be taken by a patient. Our work also aims at improving the efficiency of the proposed system.

Keywords: Cardiovascular disease; data mining; apriori algorithm

I. INTRODUCTION

Healthcare Industry (hospitals, medical centers etc) is the backbone of our Society. A major challenge that this Industry faces is providing quality services at affordable prices. Quality service implies diagnosing a patient's condition effectively, providing appropriate treatments and monitoring those treatments on a regular basis. Improper clinical decisions can lead to disastrous results. Along with diagnosing conditions and providing appropriate treatments, hospitals must also minimize the cost of clinical tests. An effective way of achieving these results is by employing appropriate computer –based information and/or decision support systems.Most hospitals today employ some sort of information systems to manage their healthcare or patient data. These systems generate huge amounts of data in the form of numbers, charts, texts images etc. Huge amount of data is stored on regular basis but unfortunately such data is rarely used to support clinical decision making. There is a large number of hidden information in these data that is largely untapped. This raises an important question "How can we turn this vast pool of data into useful information that can be used by healthcare practitioners to make proper clinical decisions?" This was the main reason for this paper.Since problem of prediction in medical dataset which involves multiple inputs can be solved by using data mining with intelligent algorithm. Artificial neural network are used nowadays for complex and difficult tasks. With the help of historical data the neural network are created with the hope that it will find out hidden dependencies and that can be used for predicting. Back-propagation is used to train feed forward neural networks and they have become a standard technique for classification and prediction tasks. The healthcare industry collects huge amounts of healthcare data and that need to be mined to discover hidden information for effective decision making. Discover of hidden patterns and relationships often go unexploited.

1.1 HEART DISEASE PREDICTION USING DATAMINING

The extraction of significant patterns from the heart disease data warehouse was presented. The heart disease data warehouse contains the screening clinical data of heart patients. Initially, the data warehouse preprocessed to make the mining process more efficient. The first stage of Association Rule used preprocessing in order to handle missing values. Later applied equal interval binning with approximate values based on medical expert advice on Pima Indian heart attack data. The significant items were calculated for all frequent patterns with the aid of the proposed approach. The frequent patterns with confidence greater than a predefined threshold were chosen and it was used in the design and development of the heart attack prediction system.Each data mining technique serves a different purpose depending on the modeling objective. The two most common modeling objectives are classification and prediction. Classification models predict under the categorical labels (discrete, unordered) while prediction models predict continuous-valued functions.

II. LITERATURE SURVEY

Numerous studies have been done that have focus on diagnosis of heart disease. They have applied different data mining techniques for diagnosis & achieved different probabilities for different methods.

Techniques	Source

1.An Intelligent Heart Disease Prediction System (IHDPS) is developed by using data mining techniques Naive Bayes, Neural Network, and Decision Trees was proposed by SellappanPalaniappan. To build this system hidden patterns and relationship between them is used. It is web-based, user friendly & expandable.	SellappanPalaniappan, RafiahAwang, *"Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008*
2.The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru. 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.	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).*

III. DATA SET

Through our proposed system, we have reduced six attributes to four which will be employed for the prediction of heart conditions. Data of multiple patients is entered into the proposed system and the diagnosed results generated by the system corresponding to those patients have been saved in the database. The resultant data thus obtained is used by the model for calculating the efficiency of the proposed system.

Original List of Attributes were as follows –

1. Type - Chest Pain Type
2. Rbp - Resting blood pressure
3. Eia - Exercise induced angina
4. Oldpk - Old peak
5. Vsl - No. of vessels colored
6. Thal -Maximum heart rate achieved

Reduced Input Attributes are as follows –

1) Blood Pressure
2) LDL – Low Density Lipoprotein (commonly known as Bad Cholesterol)
3) HDL – High Density Lipoprotein (commonly known as Good Cholesterol)
4) Triglycerides

By reducing the number of attributes as compared to earlier systems we plan on reducing the overall time required by a Doctor/ Physician to diagnose a heart condition of his/ her patient; this in turn reduces the overall cost and time required by a patient for his/her treatmen

IV. ALGORITHM

Apriori is a seminal algorithm was proposed by R. Agrawal and R. Srikant for mining frequent itemsets for the Boolean association rules mining. The name of this algorithm here is based on the fact that the algorithm uses prior knowledge of the frequent itemset properties . Apriori employs an iterative approach known as a level-wise search, where k-itemsets are used to explore (k+1) -itemsets. First, the set of frequent 1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy the minimum support. The resulting set here is denoted L1.Next, L1 is further used to find the L2, the set of frequent two itemsets, which is used to find the L3, and so on, until no more frequent k-item sets can be further found. To find each Lk requires one full scan of the database. To improve the efficiency level of frequent item sets there is an important property called the Apriori property, presented here, is used to reduce the search space. Apriori property is that All nonempty subsets of a frequent itemset must also be frequent. The Apriori property is based on the following observation. By definition, if an itemset for eg I does not satisfy the minimum support threshold i.e, min sup, then I is not frequent; that is, P (I) <min sup. If an item A is added to the itemset I, then the resulting itemset (i.e., I UA) cannot occur more frequently than I. Therefore, I UA is not frequent either; that is, P (I UA) <min sup. Thus we are using this algorithm to improve the efficiency of the itemsets. The proposed work can be further enhanced and expanded for the automation of Heart disease prediction. In the future studies that researcher can use real world data from Health care organizations and they use the available techniques for achieving higher accuracy.

V. PROPOSED SYSTEM

Our project has been mainly developed with an aim to efficiently diagnose the presence of heart disease in an individual. For this purpose we are going to use ASP.NET as our front end where in we could create a user interface to accept user details and back end would be SQL. The front end would basically work as.

- **REGISTER :** Firstly**,** if the patient is not registered or is arriving for the first time to the –doctor he should register himself so that his information can be stored in the database which would be useful in the future for diagnosis. So Initially the patient needs to register himself for the system. But, if the patient is an old user then he might go for the next step as below.

- **LOGIN:** In this step, the patient would login through his user id and access his own profile where in ASP.NET would be useful for giving access to the patient's profile.
- **USER INPUT:** After accessing the profile the doctor would enter the details of the patient as mentioned by him. The doctor would mainly undertake tests considering the four attributes in mind such as Blood Pressure – where in the values observed by the doctor would be entered in the field corresponding to Blood Pressure. Similarly all other values corresponding to the associated attributes such as LDL – Low Density Lipoprotein (commonly known as Bad Cholesterol), HDL – High Density Lipoprotein (commonly known as Good Cholesterol) and Triglycerides observed by the doctor would be entered by him respectively. Thus this would complete all the information required from the patient.
- **FINAL REPORT:** After getting the information from the patient, Data mining would be utilized where in the current details of the patient would be compared by his previous details and apriori algorithm would be used to identify if the patient has some symptoms of Heart Diseases or not. Thus, inorder to access the patients history SQL would also be used as the Back end for our System.

VI. CONCLUSION

The decision-tree algorithm is one of the most effective and efficient classification methods available. The data will judge the efficiency and correction rate of the algorithm. We used 10-fold cross validation to compute the confusion matrix of each model and then evaluate the performance by using precision, recall, F measure and ROC space. As expected, bagging algorithms, especially Apriori, showed the best performance among the tested methods. The results showed here make clinical application more accessible, which will provide great advance in healing CAD, hepatitis and diabetes. The survey is made on the decision tree algorithm Apriori towards their steps of processing data and Complexity of running data. Finally, the results are stored in the decision support repository. Since, the knowledge base is currently focused on a narrow set of diseases. And in order to improve decision support, interactions should be considered between the different medications that the patient is on.

REFERENCES

[1] P .K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules"; Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40.

[2] E.P.Ephzibah, Dr. V. Sundarapandian, "Framing Fuzzy Rules using Support Sets for Effective Heart Disease Diagnosis"; International Journal of Fuzzy Logic Systems (IJFLS) Vol.2, No.1, February 2012.

[3] A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic International Journal of Computer Applications (0975 – 8887) Volume 54– No.17, September 2012

[4] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques"; International Journal of Computer Applications (0975 – 888) Volume 47– No.10, June 2012.

[5] Mohammed Abdul Khaleel , Sateesh Kumar Pradhan , G.N.Dash; "Finding Locally Frequent Diseases Using Modified Apriori Algorithm"; International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 10, October 2013

