



Heart Disease Prediction using Hybrid Technique in Data Mining

Ankita Pimputkar¹, Prof. J. S. Dhobi²

¹Student (Master of Engineering), Computer Engineering, Government Engineering Collage, Gandhinagar, Gujarat, India

²Associate Professor, Computer Department, Government Engineering Collage, Gandhinagar, Gujarat, India

Abstract - Data mining techniques have been widely used in clinical decision supporting system for prediction and diagnosis of various disease with good accuracy. These techniques have been very effective in designing clinical support systems because of their ability to discover hidden patterns and relationships in medical data. One of the most important applications of such systems is in diagnosis of heart diseases because it is one of the leading causes of deaths all over the world. Heart disease prediction is treated as a most complicated task in the field of medical science. Thus there is a need to develop a decision support system for detecting heart disease of a patient. In this system we work on a hybrid approach for heart disease data analysis using k-means algorithm and Support Vector Machine. We propose an efficient hybrid algorithmic approach for heart disease prediction. It serves an efficient prediction technique to determine and extract the unknown knowledge of heart disease using hybrid combination of K-means clustering algorithm and SVM. To perform grouping of various attributes it uses K-means algorithm and for prediction it uses SVM. It is developed in RStudio. We get 94% accuracy.

We have seen that making cluster earlier than prediction is very useful in increasing the accuracy of prediction. We improve the accuracy as shown in result.

Keywords: Data mining, Heart Disease, K-means algorithm, Support Vector Machine

1. INTRODUCTION

1.1 Data Mining

Because of accessibility of high measure of information and a need to change over this tremendous measure of information to valuable data requires the utilization of data mining methods. Data Mining and KDD have turned out to be well known as of late.

The huge amount of data combined with the requirement for capable data analysis methods, has been defined as a data rich yet information poor circumstance. The fast developing, vast amount of information, combined and stored in huge and different data repositories, surpassed our human capacity

for cognizance without effective devices. Data gathered in expansive amount of data stores become information tombs. Critical choices are frequently made construct not just with respect to the data rich information put away in information stores, yet in addition on a leader's instinct, essentially in light of the fact that the chief does not have the apparatuses to get the profitable learning implanted in the a lot of information. Think about master framework advancements, which regularly depend on clients or area specialists to physically include information into learning bases. Sadly, this technique is inclined to inclinations and blunders, and is greatly tedious and expensive [2].

Information mining instruments perform information examination and may reveal essential information designs, contributing incredibly to business systems, learning bases, and logical research. The hole amongst information and data requires for an advancement of different information mining instruments that will transform vast vault of information into "brilliant chunks" of data and learning.

Data mining refers to fetching or "mining" knowledge from huge amount of dataset. The strategy of fetching useful knowledge and patterns in data have been given a variety of names such as data mining, knowledge extraction, data archaeology and data pattern processing but recently the terms data mining and KDD are ruling in database fields.

Medicinal Data Mining is an area of test which includes a considerable measure of misdiagnosis and vulnerability. A general system proposed for restorative information mining is appeared in fig [8].

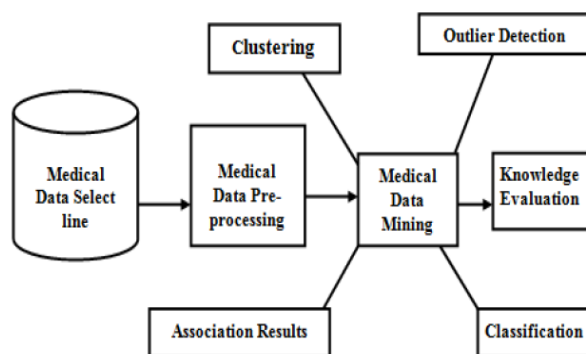


Fig -1: Framework of Medical Data Mining

1.2 Heart Disease

The heart is essential organ or part of our body. In the situation that function of heart isn't proper, it will affect the other human body parts, for example, cerebrum, kidney and so forth. Heart is empty strong organ that pumps blood through our body. The heart, blood, and veins make up the circulatory framework, which is in charge of disseminating oxygen and supplements to the body and diverting carbon dioxide and other waste items [13].

It is a pump that directs blood through the body. On the off chance that dissemination of blood in body is wasteful the organs like mind endure and if heart stops working by and large, demise happens inside minutes. Life is totally related to productive working of the heart. The term Heart malady alludes to illness of heart and vein framework inside it. It is otherwise called a Cardiovascular Disease.

CVD implies to a class of malady that include a heart or veins. It is regularly identified with atherosclerosis, a procedure whereby greasy stores ("plaques") frame in your supply routes, making them limit and conceivably square totally.

When atherosclerosis affects the significant supply routes in the body it can cause a heart assault, stroke or fringe blood vessel infection.

Heart Attack: Heart assault happens when a blood coagulation hinders the vein that provisions blood to the heart muscle itself. Angina (impermanent chest torment or inconvenience) is frequently a notice indication of a looming heart assault.

Stroke: Stroke happens when blood stream to the mind is interfered.

As the leading cause of death in the world, heart disease, according to WHO, accounts for 3.8 million and 3.4 million deaths in males and females, respectively [9].

1.2.1 Symptoms

Symptoms of Heart Disease are as follows:

- Discomfort, pressure, heaviness, or pain in the chest, arm, or below the breastbone.
- Discomfort radiating to the back, arm, throat, or jaw.
- Fullness, indigestion, or choking feeling (may feel like heartburn).
- Sweating, nausea, vomiting, or dizziness.
- Extreme weakness, anxiety, or shortness of breath.
- Rapid or irregular heartbeats

Risk factors for Heart Disease:

- Age, Angina, Blood cholesterol levels Diabetes , Diet , Genes, Hypertension, Obesity, Physical Inactivity, Smoking, Work.

1.3 Applications

1. Effectiveness of Treatment
2. Management of Healthcare
3. Disease prediction
4. Fraud detection
5. Medical Device Industry
6. Hospital Management
7. System Biology

2. RELATED WORK

1. Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors [1]

- In this paper, proposed approach focused at developing an intelligent data mining system based on genetic algorithm optimized neural network for the prediction of heart disease based on risk factors categories. It first determine the number of inputs, layer and hidden neurons of the neural network and then it uses the back propagation algorithm to train the networks using the weights optimized by GA.

2. Classification of Heart Disease using K-Nearest Neighbor and Genetic Algorithm [2]

- In this paper, proposed approach combines KNN with GA for effective classification. Proposed approach consists of two parts: 1. evaluating attributes using genetic search. 2. Building classifier and

measuring accuracy of classifier. In this approach accuracy of heart disease data is improved 5% over classification algorithm without GA.

3. Diagnosing of Heart Diseases using Average K-Nearest Neighbor Algorithm of Data mining [3]

➤ In this paper, to get improved classification accuracy and efficiency a new approach called AKNN is proposed. To remove the drawbacks and to make the KNN a faster algorithm AKNN is proposed here. In case of AKNN, super sample is created for each class which is the average of every training sample in that particular class. When the test samples are given the AKNN searches sample data and find the closest to the input. The closest neighbor is identified by measuring distance between the neighbors. The proposed approach gives higher efficiency and reduces complexity based on attribute reduction.

4. Classification and Prediction of Heart Disease Risk using Data Mining Techniques of Support Vector Machine and Artificial Neural Network [4]

➤ In this paper, the proposed approach incorporates the classes of heart disease utilizing Support Vector Machine and ANN. This paper proposes the architecture which includes Pre-processing, Preparing, Training and Testing with individual models, Evaluation of results and Prediction of heart disease risk. The proposed approach test and evaluate the projected system model in terms of specificity, precision, accuracy and sensitivity.

5. Heart Disease Prediction using Naïve Bayes [5]

➤ In this paper, the proposed approach uses Naïve Bayes Classifier. It proposed a web application that enables users to induce instant steerage on their cardiopathy through an intelligent system online. The application is fed with numerous details and also the heart disease related to those details. The application permits user to share their heart connected problems. It then processes user specific details to see for numerous health problem that might be related to it.

6. Heart Disease Prediction Using Naïve Bayes Algorithm and Laplace smoothing Technique [6]

➤ This paper approach uses classification algorithm namely Naïve Bayes algorithm and a smoothing technique Laplace smoothing. It uses details such as age, gender, fasting blood sugar, cholesterol, blood pressure etc. to predict the heart disease. Smoothing technique is used to make an approximating function that attempts to capture important patterns in the data while avoiding noise or other fine scale structures or rapid phenomena. If we end up with a probability value of zero for some value of attribute it will return a zero probability. To avoid this problems Laplace smoothing technique is used known as Laplacian correlation or Laplacian estimator.

7. Heart Disease Prediction System using Data Mining Method [7]

➤ This paper proposes a heart disease prediction system based on three different data mining techniques. 1. Naïve Bayes, 2. J48 Decision tree, 3. Random Forest. The proposed approach predict the heart disease by using medical profiles and also the performance is compared by calculation of confusion matrix. Overall system provides high performance and accuracy.

3. PROBLEM STATEMENT

Data mining technique have been widely used in clinical decision support system for prediction and diagnosis of various disease with good accuracy. As huge amount of information is produced in a

medical fields yet this information is not properly utilized. The health care system is data rich however knowledge poor. There is an absence of successful analysis methods to find connection and pattern in health care data. For this reason data mining techniques can be utilized. According to literature survey I found certain limitation for data analysis like accuracy, speed, error rate etc.

4. PROPOSED SYSTEM

In proposed system Hybrid technique is used for prediction of Heart disease. To perform grouping of various attributes it uses K-means algorithm and for prediction it uses SVM. The main aim of this system is to develop a prototype for predicting Heart disease with higher accuracy rate.

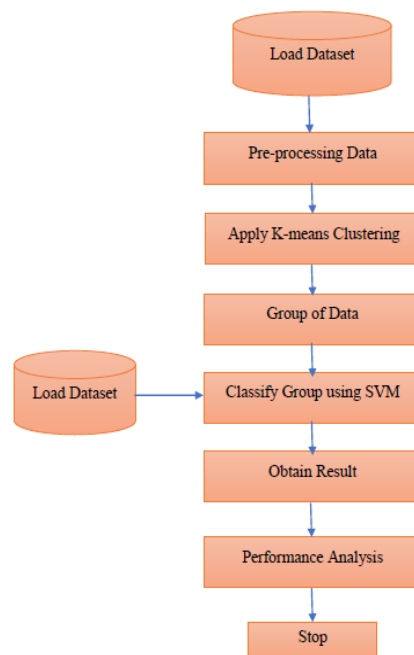


Fig -2: Flowchart of Proposed System

5. IMPLEMENTATION RESULTS

5.1 Experimental Attributes

We took online available heart disease datasets from UCI. These datasets have 76 raw attributes including predicted attribute but only 14 of them are actually most important.

Heart_statlog contains 270 records, heart_C contains 303 records, Cleveland dataset contains 304 records. We integrate all these three datasets. So we have 877 records in dataset.

Table-1: List of Attributes

NO.	ATTRIBUTES	TYPE
1.	Age	Numeric
2.	Sex	Nominal
3.	ch_pain	Nominal
4.	r_B_Pressure	Numeric
5.	Chol	Numeric

6.	f_B_sugar	Nominal
7.	r_ECG_results	Nominal
8.	maxi_heart_rate	Numeric
9.	Exercise	Nominal
10.	Oldpeak	Numeric
11.	Slope	Nominal
12.	no_of_majorvessels	Numeric
13.	Defecttype	Nominal
14.	Class	Nominal

5.2 Result Analysis

Dataset	Algorithm	Accuracy	Time(Sec.)
heartdisease(without Clustering)	SVM Linear	0.8411	19.09
heartdisease(without Clustering)	SVM Radial	0.9031	14.8
Cluster1	SVM Radial	0.98	10.19
Cluster2	SVM Radial	0.94	6.32

Fig -3: Performance Evaluation of SVM and K-means + SVM

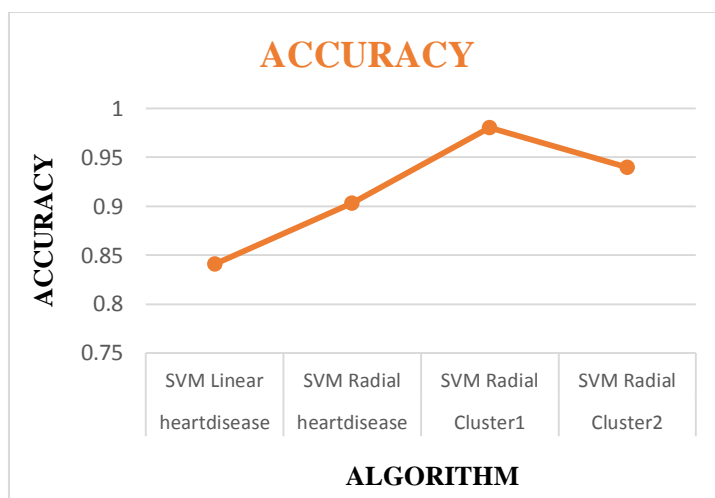


Chart-1: Accuracy graph of SVM and K-means+SVM

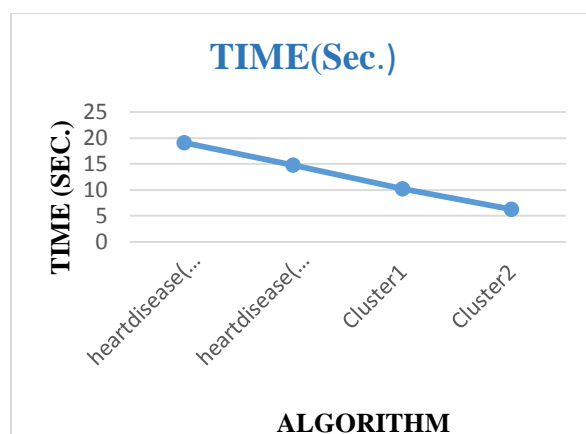


Chart-2: Elapsed Time of SVM and K-means+SVM

6. CONCLUSION AND FUTURE WORK

In this research work we have concluded that SVM algorithm combine with K-means clustering algorithm gives us faster result compare to only SVM result and gives more accuracy. In proposed system Hybrid technique is used to improve Accuracy and time complexity as compare to existing system. Here we have seen that making cluster earlier than prediction is very useful in increasing the accuracy of prediction. We improve the accuracy as shown in result.

In future, someone can use other clustering algorithm and other classification algorithm which gives better accuracy and less time than SVM algorithm. Someone can work on other of parameters and attributes of dataset to get the result.

REFERENCES

- [1] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg , “Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors” , IEEE Conference on Information and Communication Technologies, IEEE, 2013.
- [2] M. Akhil jabbar, B.L Deekshatulu, Priti Chandra, “Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm”, International Conference on Computational Intelligence: Modeling Techniques and Applications, 2013.
- [3] C. Kalaiselvi, PhD, “Diagnosis of Heart Disease Using K-Nearest Neighbor Algorithm of Data Mining”, IEEE, 2016.
- [4] Mrs S. Radhimeenakshi, “Classification and Prediction of Heart Disease Risk using Data Mining techniques of Support Vector Machine and Artificial Neural Network ”, IEEE, 2016
- [5] Garima Singh, Kiran Bagwe, Shivani Shanbhag, Shraddha Singh , Sulochana Devi, “Heart Disease Prediction using Naïve Bayes”, International Research Journal of Engineering and Technology (IRJET), March-2017.
- [6] Vincy Cherian, Bindu M. S., “ Heart Disease Prediction using Naïve Bayes Algorithm and Laplace Smoothing Technique”, International Journal of Computer Science Trends and Technology (IJCST), Mar-Apr 2017.

- [7] Ankita Dewan, Meghna Sharma, "Prediction of Heart Disease using Hybrid Technique in Data Mining Classification", IEEE, 2015.
- [8] Dr. B. Umadevi, M. Snehapriya, " A Survey on Prediction of Heart Disease Using Data Mining Techniques", , International Journal of Science and Research, April 2017.
- [9] Ramin Assari, Parham Azimi, Mohhamad Reza Tagva, "Heart Disease Diagnosis using Data Mining Techniques", International Journal of Economics & Management Sciences, 2017.
- [10] Akanksha Ahlawat, Bharati Suri, "Improving Classification in Data Mining using Hybrid algorithm", IEEE, 2016.
- [11] J. Vijayshree, N.Ch.SrimanNarayanaIyengar, "Heart Disease Prediction System using Data Mining and Hybrid Intelligent Techniques: Review", International Journal of Bio-Science and Bio-Technology, 2016.
- [12] K. Rajlakshmi, Dr. S. S. Denakaran, N. Roobini, "Comparative Analysis of K-means Algorithm in Disease Prediction", International Journal of Science, Engineering and Technology Research, July 2015.
- [13] Sonam Nikhar, A. M. Karandikar, "Prediction of Heart Disease using Data Mining Techniques", International Journal of Engineering and Technology, Feb 2016.