



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH


IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 1, January 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)



# Case Study on Prediction of Heart Disease Using Data Mining Techniques

Alan Rachel George<sup>1</sup>

P.G. Student, Department of Computer Science, Kristu Jayanti College (Autonomous), Karnataka, India

**ABSTRACT:**The human heart is an organ that siphons blood all through the body by methods for the circulatory system, giving oxygen and enhancements to the tissues and taking out carbon dioxide and diverse squanders. Coronary sickness is one of the most savage issues in the whole world, which can't be seen with an independent eye and arrives instantly when its requirements are reached. Clinical administrations evaluate the datasets regularly by professional learning and action. Data Mining techniques are utilized to remove helpful data and bits of knowledge from this information. The outcomes might in the appropriate treatment just as sparing existences of many. Data mining algorithms, for example, KNN, Naïve Bayes, and ANNs can be used to foretell the presence of coronary illness and guarantee appropriate treatment heretofore. The aim of this research paper is to conduct a case study to predict the chances for a person to get heart illness using different data mining techniques.

**KEYWORDS:**Data Mining techniques, Heart disease prediction, Computer-Aided Support System

## I. INTRODUCTION

"Data mining is the method of finding important details an immense measure of information in the database". The information sources can encompass databases, records warehouses, the web, and different facts repositories or a record which are streamed into the scheme dynamically.

Coronary, a muscular organ that pumps the blood throughout the circulatory manner by rhythmic contraction and dilation. In the circulatory system, blood is transferred throughout the veins in the heart. This muscular method performs a crucial position because it transports oxygen, blood and different supplies to the variety of physique components. It can trigger critical wellbeing situations together with loss of life if the guts would not toil appropriately. It ends up in several illness, incapacity and death. The period of coronary illness is every so often used interchangeably with the time period cardiovascular disease. A heart problem most often refers to situations that contain narrowed or blocked blood vessels that may result in a coronary attack, chest ache or stroke. Modifiable danger elements incorporate weight, smoking, diabetes, high cholesterol, High blood pressure absence of bodily movement, etc. sickness prognosis performs a number one position in the medical field.

Data mining algorithms deal with the difficulty of medical dataset prediction involving several inputs. The selection of programs for the prediction of various ailments are proposed and carried out by utilizing completely programs and methods. This paper predicts lots of coronary sickness prediction by using the approaches of information mining proposed in the latest years.

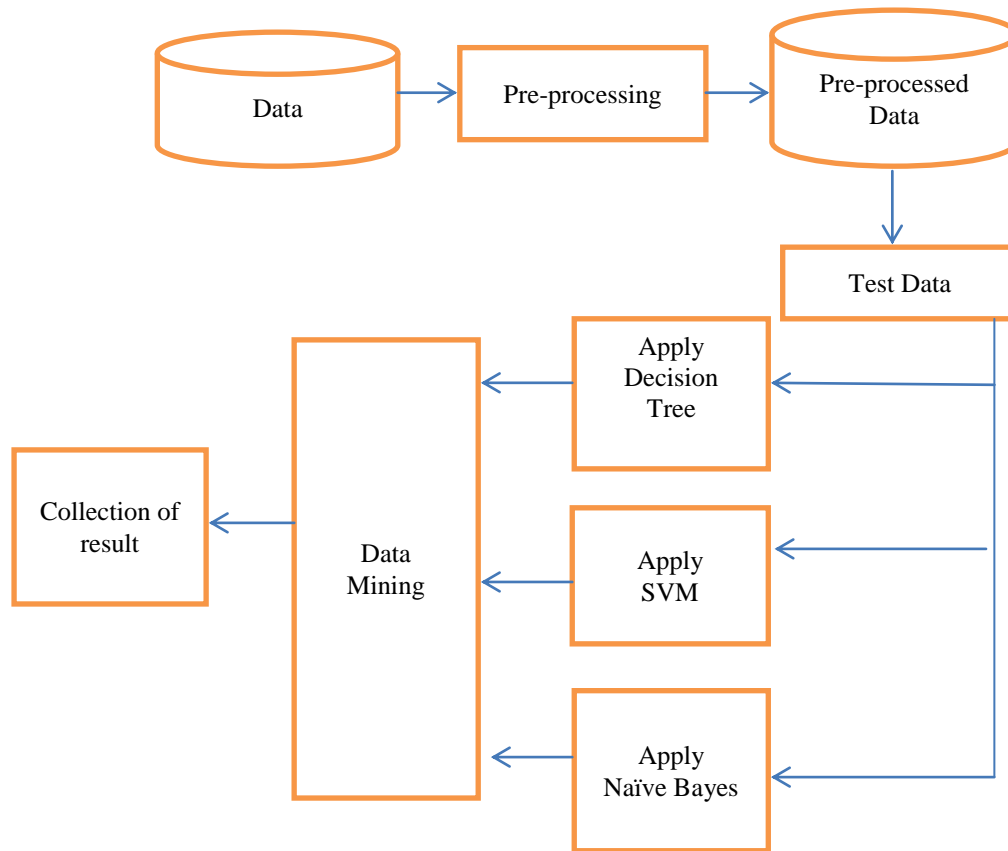


Fig.1 Flow chart of Processing the Data

## II. RELATED WORK

- G. Purusothaman and P. Krishnakumari et al [1] proposed an audit on information mining techniques for prediction of disorder . Various methods are applied to guage heart condition risk prediction. By using classification techniques, heart condition prediction model find results a minimum of expense and exertion. Following are the 2 prediction models used for data analysis. Applying single model to voluminous model associated with heart data is first one and second is for implementing combined model to urge hybrid model.
- K.Polarajuet al [2], suggested disorder Forecasting using several rectilinear regression , and K shows that multiple correlation analysis is acceptable for forecasting risk of cardiac attack . The analysis is completed by means of designing data collection consisting of 3000 instances with 13 specific attributes previously specified. the gathering of knowledge is separated into 2 parts, which account for 70% of the info used for teaching also 30% for research. Depending resting on the outcomes, it's obvious to Regression method's arrangement accuracy is best related to other methodologies.
- Vincy Cherian et al [3] proposed prediction using Laplace smoothing and Naïve Bayesian algorithms. The prediction model is assumed to stay away irrelevant diagnosis test conducted and delay in starting right treatment .Therefore, by early diagnosing patient with heart condition both time and money are often saved. Therefore, doctors can easily diagnose without unnecessary treatment provoked due to inexperience and intuitions made by doctor. This paper additionally looks at the exactness when attributes are decreased for prediction. A voluminous training dataset would definitely provide appropriate results which may train the system model properly. The framework are often extended to include different fields for various ailment expectation.

- M. A. Jabbar et al [4] proposed Association Rule mining supported the sequence number and clustering for attack prediction. The whole database is split into partitions of equal size. The dataset with fourteen attributes was utilized in that employment and also each cluster is taken into account one at a time for calculating frequent item sets. This approach reduces main memory requirement. To predict the goods attack in an efficient way the patterns are extracted from the database with significant weight calculation.
- Mai Shouman et al [5], proposed k-means clustering with the selection tree method to predict the center disease. In their work they suggested several centroid selection methods for k-means clustering to increase efficiency. The thirteen input attributes were collected from Cleveland Clinic Foundation heart disease data set. The specificity, and exactness are calculated with different initial centroids selection methods and different numbers of clusters.
- Tsien et al [6], in their study indicated that classification trees, which have certain advantages over logistic regression models, may perform similar to logistic regression models in the diagnosis of patients with MI.
- Kiruthika Devi, D. Kalita et al [7] proposed in health care domain to survey numerous data processing tools for heart condition prediction and diagnosis. A coronary illness expectation model uses algorithms of knowledge mining to support doctors in predicting disorder result on the idea of clinical information. within the field of medicinal services, data processing classification techniques like SVM, Naïve Bayes, decision tree etc. helps in better deciding . Algorithms gathering assist to form decisions more accurate and quicker.
- S. B. Patil and Y. S. Kumaraswamy et al [8] proposed a MAFIA (MAximal Frequent Itemset Algorithm) to extract the data relevant to attack from the warehouse. Then the many weightage of the frequent patterns are calculated. Further, the patterns significant to attack prediction are chosen supported the calculated significant weightage. These significant patterns are often utilized in the event of attack prediction system.

### III. DATA MINING ALGORITHMS

In every field of education, we need prior knowledge to understand and analyse that field very well, prior knowledge becomes base for successful understanding and analyses of any study. So before we start to study the actual content of this paper we have to study and understand the basic concepts related to the paper that will help us to understand and comprehend the paper very well. Various algorithms formulates due to different research works on data mining. These techniques are straightforwardly utilized for developing frameworks or to find crucial inferences and conclusions from the resulted dataset. Various well-known techniques are Decision Tree, SVM (Support Vector Machine), Naïve Bayes, etc are discussed.

*a) Decision Tree:* This tool utilizes a tree-like model or graph of selections so that possible results which include accident outcomes and utility. This tool is one among the common approach to point out algorithms. Decision tree uses the tree demonstration to crack the problem during which every leaf node correspond towards the category tag and aspects be represented on the inner node of the tree. It takes as input a record or object described by a group of attributes and returns a "decision with predicted output value for the input". The input attributes are often discrete or continuous. One by one mapping from root nodes to the leaf node so that this will effortlessly be changed within the sort of rules and eventually by following these norms, appropriate results are often found. Random Forest and Logistic Tree Model are Decision tree implementation algorithms.

*b) Support Vector Machine(SVM):* SVM is employed for accessing information and pattern for multivariate analysis and classification. The objective is to seek out most accurate classification function to acknowledge individuals from the 2 classes within the training datasets. It works well on data sets that have many attributes and ensures that the simplest such function is found by maximizing the margin between the 2 classes. It differentiates data by searching the simplest hyperplane that divides all data points of 1 class from those of the opposite class. The optimal SVM is to elucidate the specimens as spatial data, intended to divide the specimens within the different groupings by a good distance that's as varied as probable. So fresh conditions are built for the actual place and are forced to contribute to a corporation centered on the side of the slit they strike.

*c) Naïve Bayes Algorithm:* A Naive Bayes classifier may be a supervised machine learning technique supported Bayes theorem. In simple terms, a Naive Bayes classifier assumes that the presence or absence of a specific attribute of a category is independent to the presence or absence of the other attribute of that class. Since the 1960s, Naïve Bayes former regarded expansively. It had been introduced within the information extraction group within the early 1960s and



remains a baseline strategy for assessing papers as belonging to at least one of the previous group with word frequencies since identifiers. Bayes theorem is calculated as follows:  $P(C|X) = P(X|C) * P(C) / P(X)$ , where C is that the class such P(X) is constant for all classes and X is data tuple.

IV. EXPERIMENTAL RESULTS

In this investigation, the precision of three Data mining methods is looked at. The objective is to have high exactness, other than high accuracy and review measurements. Despite the fact that these measurements are utilized all the more regularly in the field of data recovery, here we have considered them as they are identified with the other existing measurements, for example, particularity and affectability.

The heart failure clinical records data set was downloaded from UCI repository. The data set consist of 300 rows and 13 attributes

Figures show the results of prediction done in Rapid Miner. Fig 2 (a) shows the result of prediction done by using Decision tree. Fig 2 (b) shows the result of prediction done by using Naive Bayes. Fig 2 (c) shows the result of prediction done by using SVM.

Fig 2 (a)

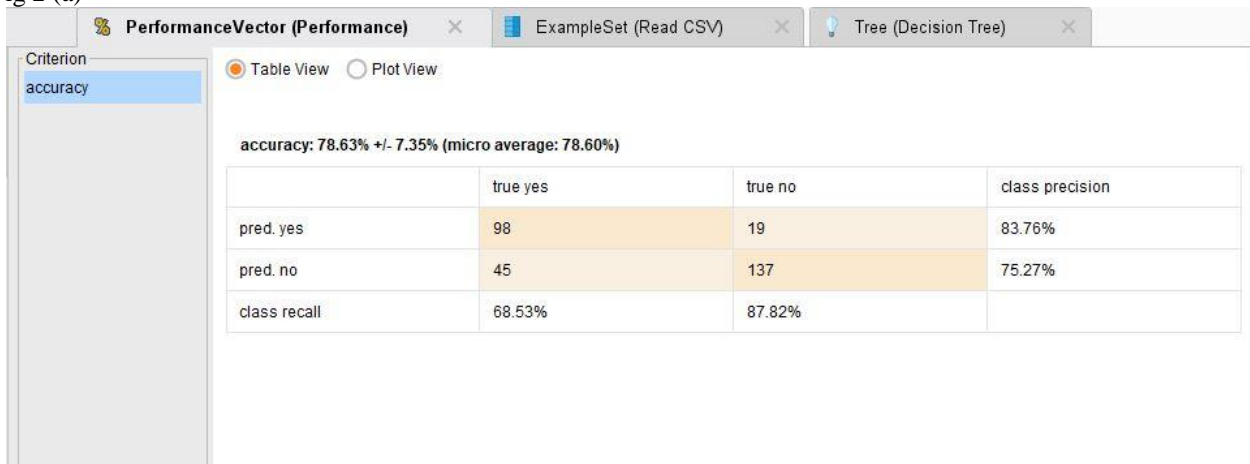


Fig 2 (b)

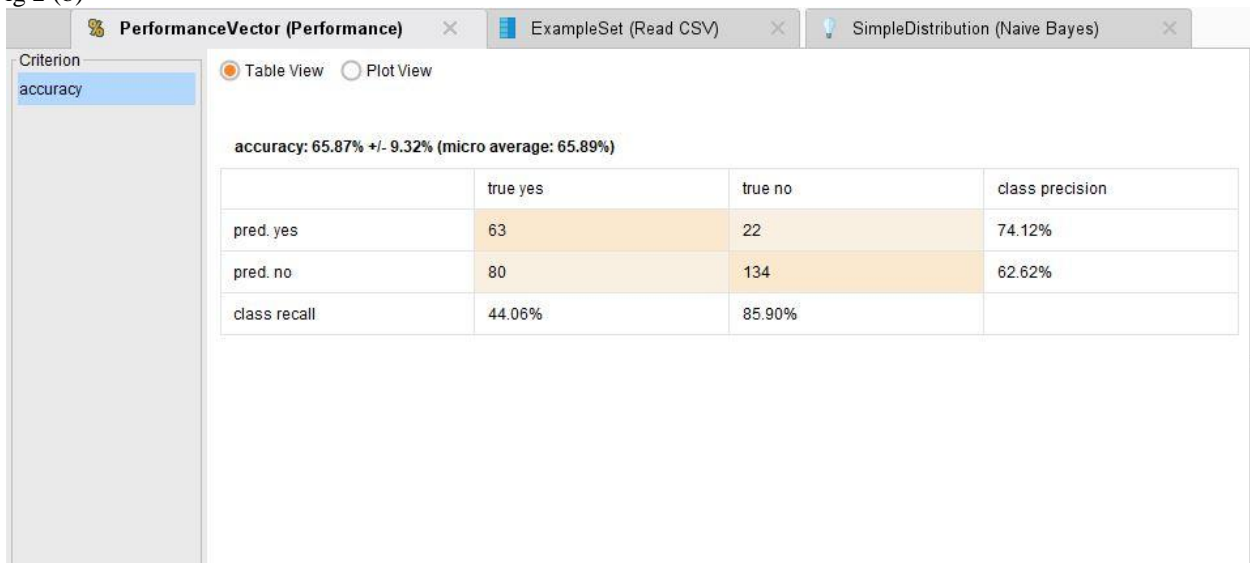
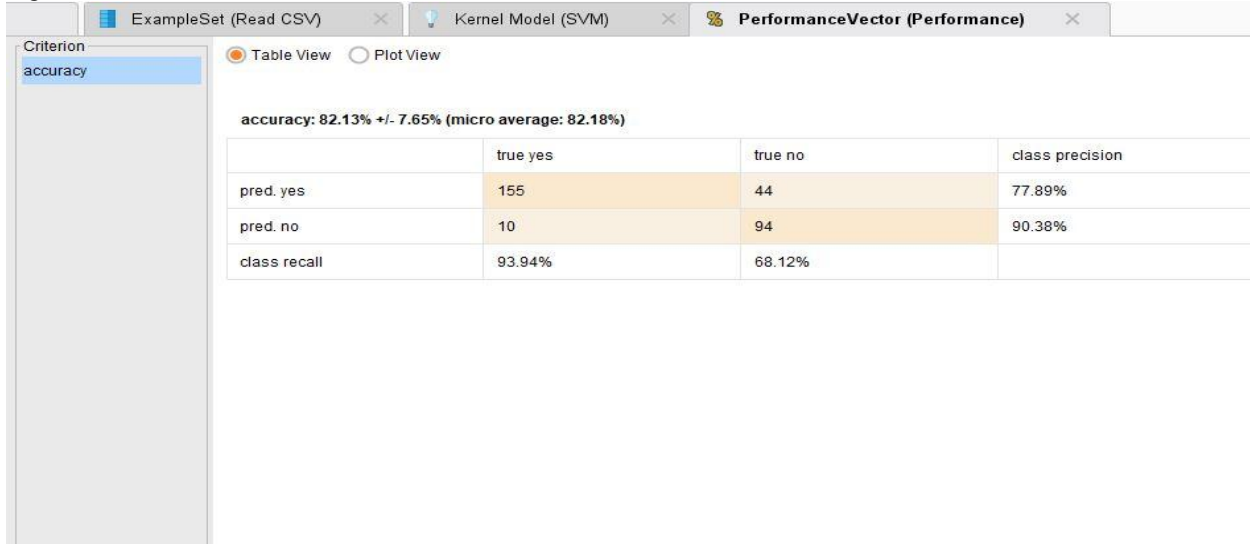




Fig 2 (c)



From the investigation of three Data mining methods i.e. , Decision Tree, SVM (Support Vector Machine), Naïve Bayes, the algorithm which gave the most accurate result was SVM with 82.12% accuracy.

### V. CONCLUSION

Over the planet huge reason for death is heart diseases. On the opposite hand, their early diagnosis helps in improving patients’ health status and leads to life-saver. This paper investigates different experiments from various surveys that are conducted to seek out the foremost accurate classifiers in order that heart diseases are often predicted as soon as possible

### REFERENCES

- [1] G. Purusothaman\* and P. Krishnakumari,” A Survey of Data Mining Techniques on Risk Prediction: Heart Disease”, Indian Journal of Science and Technology, Vol 8(12), DOI: 10.17485/ijst/2015/v8i12/58385, June 2015 ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645
- [2] K. Polaraju, D. Durga Prasad, “Prediction of Heart Disease using Multiple Linear Regression Model”, International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.
- [3] Vincy Cherian Bindu M.S, “Heart Disease Prediction Using Naive Bayes Algorithm and Laplace Smoothing Technique”, ISSN: 2347-8578 Vol. 5(2), PP. 70-71, 2017.
- [4] M A Jabbar,” CLUSTER BASED ASSOCIATION RULE MINING FOR HEART ATTACK PREDICTION” Journal of Theoretical and Applied Information Technology 31st October 2011. Vol. 32 No.2
- [5] Shouman M, Turner T, Stocker R. Integrating decision tree and K-means clustering with different initial centroid selection methods in the diagnosis of heart disease patients. Proceedings of the International Conference on Data Mining; 2012.
- [6] C. L. Tsien, H. S. F. Fraser, W. J. Long, and R. L. Kennedy, “Using classification trees and logistic regression methods to diagnose myocardial infraction,” in Proc. 9th World Congr. Med. Inf., vol. 52, pp. 493–497, 1998
- [7] S. Kiruthika Devi, S. Krishnapriya and D. Kalita, "Prediction of Heart Disease using Data Mining Techniques," Indian Journal of Science and Technology, vol. 9, no. 39, pp. 1-5, 2016.
- [8] S. B. Patil and Y. S. Kumaraswamy, “Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction,” International Journal of Computer Science and Network Security (IJCSNS), vol. 9, no. 2, pp. 228–235, 2009.



**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor

**Impact Factor:**  
**7.512**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details