

# A review paper on the Prediction of data mining model on heart disease

Sonam kashyap<sup>1</sup> and kalpana yadav<sup>1</sup>

Information technology Dept. of Indira Gandhi Delhi Technical University For Women

*Abstract – Prediction of heart disease is based on medical practitioners on how accurately they can predict a heart disease in a patient. There are many attributes that changes the state of heart worstly. Data mining is a technique where a large data set is being analyzed by a number of tools to find the important information which is not extracted previously. After the knowledge has been discovered, it is used to precisely predict data mining or heart disease. Problems related to heart disease are the major problems in health field in this modern era. This paper gives a brief description about prediction of heart related diseases by the use of data mining techniques. The major reason to use data mining technologies is increasing amount of data in healthcare sector which is “not minded”. Heart Disease Prediction Model supports various medical executives in guessing heart disease chances established on the basis of data of patients provided by various clinics. A organ known as heart have the most leading role in a body of human, if it is not able to work properly then a number of deaths occur that’s why there is a need to predict the main causes and symptoms of heart disease.*

*Keywords –Heart disease, Data mining, Naïve Bayes, Decision Tree, Random Forest, logistic Regression.*

## I. INTRODUCTION

Concerning security of information is not all about the unauthorized access by an attacker. Information Security is the operation of avert illegal accessibility, revelation, interference, moderation. Statistical data can be physically stored or can also be stored on electrical devices. Types of information like details provided during online banking or your description of profile on social media, feeded data in mobile phones etc. because of the many applications of security of information like encryption and decryption of data, cyber forensics etc. Data mining help researchers to find large amounts of data to determine patterns that was get “lost in the noise.” Beginning of data mining is started from raw data, which can be ranges from numeric observations to complex examination with thousands of variables. Data mining methods uses some particular estimating procedures to find out meaningful explications and useful patterns in the sensible data. Students who wants to be a data mining scholar have to study the tools and equipment of data mining so that they can improve the concepts of analysis and processing of large and complex datasets. There is a strict need to adopt data mining techniques in the field of medical practitioners so that they are able to easily diagnose heart patients. There are some applications of data analysis that include treatment centres analysis aimed to improve treatment policies and prevention of errors in hospitals, early diagnosis of diseases, prevention of diseases and reduction of deaths in a hospital that a data mining scholar have to learn.

## II. DEFINING INFORMATION SECURITY

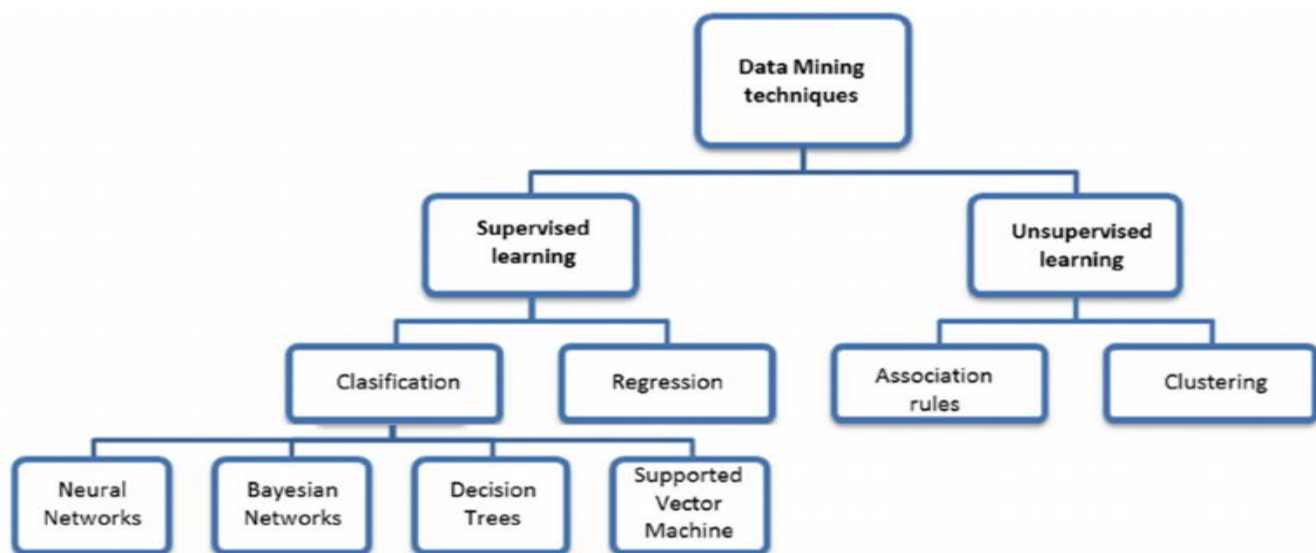
Information security is the application and practice of allowing access to information to people who are the authorized person to see that information. Authorization of information includes access to information stored in a computer, enclose all the information under the control of an organization authority. The concept of authentication and authorization was invented by U.S. Central Intelligence Agency who wants to make sure that documents were safe from being modified or not accessed by people who were not allowed to obtain them, especially that information which was classified in different department. Information security doesn't only deal with the data stores on computer by every organization. Information security is firstly a management situation. A best information security deals with the top and move towards a good IT policy. Everyone should have a duty to make sure information should be secure and reliable for both public and private.

### III. DEFINING HEART DISEASE

The topic heart disease itself contains a range of conditions that influence a human body easily. The term heart disease can be swap with the term cardiovascular disease. CVD refers to the condition which holds insufficient or blocked blood vessels that is conduct to a typical heart attack, angina or may be a stroke. The heart is an important organ of a human body. Whom main function is to pump the blood through the body. If the movement of blood through the body is blocked due to some major problems like brain damage and the heart is not working properly due to such problems, death occurs within some seconds. Heart disease define as the disease which are related to heart and vessels and also values are not functioning properly in a heart. There are many symptoms other than heart related symptoms which can help in detecting the heart disease in a normal patient before that patient can be diseased at his heart. That symptoms can be family history, obesity, tobacco, smoking, unhealthy food etc. symptoms like them are used to find patients with heart disease. In identification of a disease, doctors need patients current results of test and their past references.

### IV. DATA MINING ALGORITHMS

Data Mining is a method of insignificantly discovering the non-findable data in the past and potential major data[1]. Research work like mathematics, genetics, exploration and prediction-classification are comes under the field of data mining. Diagnosis is a complex task and needed a highly experience doctors and it also demand to executed accurately. Data mining prediction model help doctors to predict a large amount of patient data in few minutes instead manually it takes a long time to predict. By using some data mining algorithms we can predict heart disease patient's record. In this paper our main agenda is to provide a proper data mining algorithms details and a literature review. various data mining techniques are represented in fig.1



#### A. Naïve Bayes:

Naïve bayes is one of the classification method of data mining which is construct on bayes theorem with the expectation that predictors are unrelated to each other. Naïve Bayes classifiers are trained in a regular way in supervised learning technique. Naïve bayes model is beneficial for extremely big datasets.

Bayes Rule:

$$P(O|N) = (P(N|O).P(O))/P(N)$$

Where,

$P(O|N)$  = Posterior Probability

$P(N|O)$  = Likelihood

$P(O)$  = Class Prior Probability

$P(N)$  = Predictor Prior Probability

probably Naïve bayes defines determinant and minor relationship between two randomly generated events. It helps to assess posterior possibilities that holds observations. Naive Bayes compute the probability which is showing the correct or incorrect result in diagnosis.

#### B. Decision Tree:

Decision Trees (DTs) are the part of previously provide trained data which can also be known as supervised method of learning whose nature is non parametric and used for data mining classification techniques[3]. Decision trees are appear as trees. Decision tree has tree structured classifier and are of two forms of nodes: nodes which are used for decisions and Leaf nodes. Decision node decide which direction we have to go for taking a decision. Leaf node indicates the classification of the example. The major property of decision tree is that it can be used for both classification and regression. The advantages of decision tree are easy understanding of datasets and can handle both numerical and unambiguous data. The limitation of decision tree is it works only and only on very small criteria of trained data and if the data get complex accuracy of decision tree may get affected.

#### C. random forest:

Random forest is a grouped classification method that consists of multiple decision trees. It is the technique which is more accurate than any other machine learning algorithms. Good accomplishment on thousand number of difficulties because of its nature which is non-disturbable to noises produce in data set and also not influenced by over fitting. This technique is built by merging the outcomes of multiple decision trees, each of which are isolated trained

#### D. Logistic Regression:

Logistic regression a term of regression which is utilize to estimate binary or multi- category variable and also the response variable is separate. It's employed to classify low dimensional knowledge having non-linear boundaries. It offer changes within the share of variables and provides the rank to variable individually as per their importance. The major thought come in mind is to supply regression to find out the fulfillment of every independent variable precisely.

## V. LITERATURE REVIEW

S.NO	AUTHORS	YEAR	TITLE	DESCRIPTION
1	G.Subbalaxmi et al	2011	Decision support in heart disease prediction system using naïve bayes	They develop a decision support system for predictable heart disease methodology which is executed with a questionnaire software which is web based. They suggest that their research can help to train doctors, medical practitioners and researchers to provide better treatment to patients which are affected by heart disease. Their queries are `what if` queries which are hard to be answered using traditional depository of datasets.
2	M.Akhil jabbar et al	2013	Classification of heart disease using k-nearest neighbour and genetic algorithm	They gives their detailed review of k-nearest neighbor and used this data mining methodology to guess heart disease chances. They give a brief approach to combine KNN and genetic algorithm to improve the performance and to decrease the attributes of the dataset. This research recommended a new algorithm which is a hybrid algorithm with the smallest steps in it. They test their proposed method with 6 medical dataset and 1 non-

				medical data set which are chosen from UCI repository.
3.	Chaitrali S Dangare et al	2012	Improved study of heart disease prediction system using data mining classification techniques	They extend the research by adding two more attributes in 13 attributes which are previously used i.e. obesity and smoking. Their result shows the accuracy of various data mining algorithms like neural network, decision tree and naïve bayes as 100, 99.62 & 90.74 in percentage respectively. They also shows the structure of multilayer perceptron neural network.
4.	Mai Shouman et al	2012	Using data mining techniques in heart disease diagnosis and treatment	Their stocker aims to apply the hybrid algorithm which is further in working to show more accurate prediction results during the diagnosis of heart disease patients. Their paper aims is to describe and develop a tool which can be implant in hospitals control system to give advice to health care professionals by providing suitable treatment for heart influenced patients. They suggest that hospitals are not able to provide the quality of services instead of that they provide the same types of services. They use various data mining applications like better procedure making and elimination of errors produce during testing in hospitals and early detection of heart disease.
5.	Sellappan Palaniappan et al	2008	Intelligent heart disease prediction system using data mining techniques	They use systems which are intelligent in predicting the heart problems and are able to handle complicated queries which were not given by traditional managements systems. This paper uses .NET Platform for its user friendly and reliable nature. They describe their prediction model using charts and matrices methods to evaluate its effectiveness in intelligent systems. In their research they conclude that the best method to predict heart disease in the naïve bayes model. They also suggest that their research can be further expanded like combining data mining and text mining or can be using continuous data instead of categorical data.

6.	Yanwei Xing et al	2007	Combination of data mining methods with new medical data to predict outcome of coronary heart disease.	Their aims to develop such data exploration algorithm which is used to predict the endurance of heart problems. They perform a manual observation with a 6 month review of 1000 CAD cases. To measure the impartial estimation of prediction model they use for fold cross validation technique for the purpose of performance comparison of the three algorithm that they used in this paper. In their result best prediction with the accuracy of 92.1% is the support vector machine (svm) and then artificial neural network came out with the accuracy of 91.0% and the worst algorithm out of the three algorithm is decision tree which gives accuracy of 89.6.
7.	Singh Navdeep et al	2018	Heart disease prediction system using hybrid technique of data mining algorithms	They style a perceptive method for influenced heart acknowledgement using data mining and exploration methods that are fit for enhancing the consistency of heart infections conclusion In this paper the researches used data sets of three hundred three records and fourteen elements that are composed from the net dataset depository of Archive.ics.edu/ml/datasetFor the execution of the project, the platform used is Python 3.6. They give a model that use two different classification data mining algorithms i.e. genetic rule and Naive Bayes.
8.	B.Venkatalakshmi et al	2014	Heart Disease Diagnosis Using Predictive Data mining	They examines that after conducting many experiments on the same dataset they conclude that naïve bayes algorithm gives better performance than decision tree. This paper also infers that genetic algorithm can lessen the size of data to give the minimal subgroup of elements that are used for heart disease prediction in future.

9.	Aqueel Ahmed et al	2012	Data Mining Techniques to Find Out Heart Diseases	They defines various classification algorithm and gives their performances. In this research algorithms like decision tree and SVM give much more accurate results than any other methodology. They suggest many different measures which are used to guess heart disease chances in patients that can help in future for medical participators in finding more accurate data.
10.	Nidhi Bhatia et al	2012	An Analysis of Heart Disease Prediction using Different Data Mining Techniques	They inspect data mining classifiers and explained them in their work that are appear in recent years for predicting cardiovascular disease designation. This paper concludes that decision tree with the assistance of genetic algorithm program shows better accuracy and performance and gives set choices. In their analysis, neural network gave the highest accuracy of 100%. The target of their effort supplies a brief study of different data processing methods.
11.	Abhishek taneja et al	2013	Heart disease prediction system using data mining techniques	In this paper, the researcher is discuss about evolve a cost effectual treatment which uses mining of data techniques for ease of database DSS. They use treatment of heart disease using its symptoms as a composite activity. This paper diagnosis heart related problems by using different data mining techniques or algorithms. A complicated transthoracic echocardiographic report data set is used to diagnosis heart disease by using ECG .After comparison of data mining algorithm they get some result and the most effective algorithm is J48 classifier with 95.56% accuracy rate.

12.	G. Purusothaman et al	2015	A Survey of Data Mining Techniques on Risk Prediction: Heart Disease	In this paper, the comparison of different algorithms are used to determine the best accuracy given algorithm. In this researcher is discuss about hybrid model which is the combination of two algorithms and they were also discuss about the heart disease is time precious, for saving the life of many patients time effective risk identification is very much needed.
-----	-----------------------	------	--	--

## VII. CONCLUSION

In the conclusion of this paper we want to conclude that in the recent era Heart risks and problems are one of the major causes of deaths in this world and the researchers which are in the field of heart disease explorations need a early prediction of heart disease. The systems assigned in computers of hospitals needs a heart disease prediction system which helps various data scientists who are in the field of heart specialization for heart disease diagnosis. Four types of Heart Disease classification algorithms are reviewed in this paper along with some literature review of papers which are previously known for heart disease prediction. By the analysis of various papers it is concluded that, data mining have a greater role in classification of heart disease and its various symptoms which are known as attributes. This paper can be used for finding the accuracies of different data mining algorithms by using various tools of analysis of data mining classification system.

## VIII. REFERENCES

### Journal References

- [1] Ishtake S.H ,Prof. Sanap S.A., “Intelligent Heart Disease Prediction System Using Data Mining Techniques”, International Journal of Healthcare & Biomedical Research,2013.
- [2] Monali Dey, Siddharth Swarup Rautaray, “Study and Analysis of Data mining Algorithms for Healthcare Decision Support System”, International Journal of Computer Science and Information Technologies, 2014.
- [3] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin,“A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining”, International Journal of Applied Engineering Research, 2012.
- [4] Purushottam , Prof. (Dr.) Kanak Saxena, Richa Sharma, “Efficient Heart Disease Prediction System”, Elsevier B.V.,2016.
- [5] Sharma Purushottam, Dr Kanak Saxena, Richa Sharma” efficient Heart Disease Prediction System using Decision Tree” , IEEE International Conference on Computing Communication and Automation (ICCCA-2015),May 2015.
- [6] V. Krishnaiah, G. Narsimha, and N. Subhash Chandra,” Heart Disease Prediction System Using Data Mining Technique by Fuzzy K-NN Approach”, Springer,2015.
- [7] Chaitrali S. Dangare,” Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques”, International Journal of Computer Applications (IJCA), June 2012.
- [8] Dilip Roy Chowdhury, Mridula Chatterjee R. K. Samanta,“An Artificial Neural Network Model for Neonatal Disease Diagnosis”, International Journal of Artificial Intelligence and Expert Systems (IJAE),2011.
- [9] K. Srinivas, G. Raghavendra Rao and A. Govardhan, “Survey on prediction of heart morbidity using data mining techniques”, International Journal of Data Mining & Knowledge Management Process, 2011.
- [10] K. Usha Rani, “Analysis of heart diseases dataset using neural network approach”, International Journal of Data Mining and Knowledge Management Process , 2011.

### Book References

- [11] Jiawei Han and Micheline Kamber – Data Mining – *Concept and Techniques*; Second Edition; Elsevier Inc, 2006.