

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES RULE BASED APRIORI APPROACH FOR HEART DISEASE PREDICTION

Pawandeep Kaur*¹ & Dr. Gaurav Gupta²

*¹Student, Dept. of Computer Science and Engineering, Punjabi University, Patiala, India

²Assistant Professor, Dept. of Computer Science and Engineering, Punjabi University, Patiala, India

ABSTRACT

The work on diverse forms of data-mining techniques is used for the prediction of severe problems of the heart disease. It is usually done by the use of different data-mining tools. Apriori Rule base is useful in many domains playing a prime role in clinical research methods where the algorithms for Apriori Rule base processes are preferred to extract the knowledge i.e. hidden in the data-set. The scientifically discovered patterns helps in aiding the process of decision making which helps in saving the human lives facing the problem of heart disease. Health care organizations are undergoing major challenges leading to catastrophic results that are unacceptable in nature. The bad decisions related to health may cause death of human. In order achieve a cost-effective and corrective treatment, a good methodology should be adopted. The traditional methods of medical system reveals failed results and does not reveal the actual cause of pain. The main motive of this thesis is to design a system that helps in determining and extracting the knowledge i.e. hidden from the past patient's records. It further detects the heart disease and allows the practitioners to take accurate decisions. To overcome these challenges decision support system and computer based information system must be used for better precise results. The process of heart abnormality prediction involves the use of a distinct form of technique known as Apriori Rule base. Relying on data-mining techniques, the process of mining has dedicated its career in planning of how to draw conclusions from diverse form of information.

Keywords: datamining, rule, Apriori, Classification algorithms

I. INTRODUCTION

The design of this paper comprises the work on diverse forms of data-mining techniques used for the prediction of severe problems of the heart disease. It is usually done by the use of different data-mining tools. Data Mining is useful in many domains playing a prime role in clinical research methods where the algorithms for data mining processes are preferred to extract the knowledge i.e. hidden in the data-set. The scientifically discovered patterns helps in aiding the process of decision making there saving the human lives facing the heart disease problem. Health care organizations are undergoing major challenges leading to catastrophic results that are unacceptable in nature [1]. The traditional methods of medical system reveals failed results and does not reveal the actual cause of pain. Data mining represents a methodology of data separation from large arrangements of databases. The aim of the data mining strategy is to extract the valuable information from the composed record of data [2]. Data Mining is the procedure that uses an assortment of data analysis and modelling strategies to discover patterns and connections in data that might be utilized to make precise expectations. These are very commonly used in medical research as well as the engineering application analysis [1]. There are different methods of data mining like classification, clustering, association rule mining and so on. Every strategy has its own significance as indicated by his part. There are different uses of data mining in different fields like education, scientific and engineering, healthcare, business and some more. It can help you to choose the correct prospects on whom to focus [4]. To overcome these challenges decision support system and computer based information system must be used for better precise results.

The process of data-mining involves the following phases as represented below in figure.1

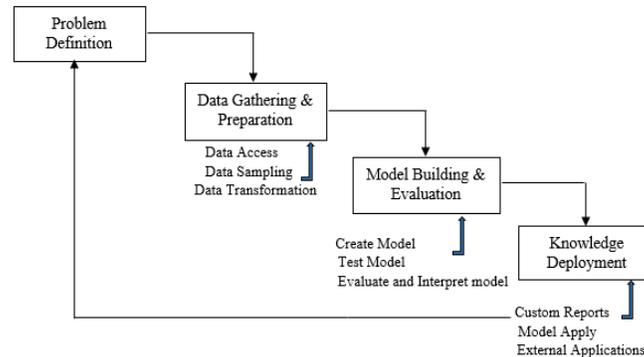


Fig.1 Modern Data Mining

- **Problem definition:**The project of data-mining starts with the experiencing the problem. The experts of different fields such as data-mining, domain and business expert jointly work together to define the main objectives and the basic requirements of the system. Then the objective of the system is transformed into problem definition of data-mining. Here, the tools are not required.
- **Data preparation:** In this process the data is collected, sampled, transformed, and prepared for the process of data-modelling. The domain-experts builds the data-model by collecting, cleansing and formatting the data. In the phase of multiple times. The typical task of data preparation comprises of distinct tool modelling such as selection of attributes, tables, and records. Here, the meaning of the data is not changed at all.
- **Modelling and Evaluation:** Firstly, in the phase of modelling, the data model is created where both the evaluation phase and the modelling phase are jointly coupled. The parameters value selected can be changed repeatedly until an optimized solution is achieved resulting in a high quality model formation. Secondly, the data-mining experts evaluates the model i.e. created. If the model does not fulfil the expectation, the process goes back gain to the phase of modelling and rebuilds the model again by changing their optimized values that was obtained earlier. If the model is satisfied, the experts can extract the needed explanations and evaluates the following questions:
 - Does the created-model achieve the objective?
 - Have all the issues been considered or not?

In the end, the mining experts use the results in different ways.

- **Deployment:** In this process, the mining experts of the data use the results of mining by transporting the results into data-base tables or other forms of applications as required.

A. *Apriori Rule base*

The operation of Apriori Rule base includes various methods of separating the useful data and to acquire precise knowledge from massive arrangements of data. The objective of this methodology is to wrench out beneficial information from the composed or collected records of data. In early stages it merges statistical analysis, database and machine learning which enables to search the unknown trends and patterns in the databases thereby using the extracted information to assemble deductive models [10]. Apriori Rule base is a procedure of using an array of data and different modelling strategies to discover diverse patterns and the interconnection of the links among data empowering a precise form of expected information with distinct behaviour. The following flowchart represents the work-flow of the Apriori Rule baseProcess.

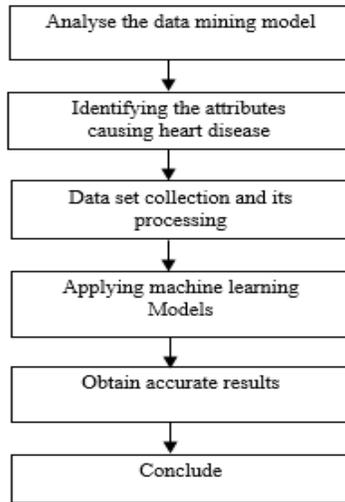


Fig.2 Work Flow: Apriori Rule base Process

The analysis of data, discovery, and building model strategies is commonly an iterative process as it targets and recognizes the contrasting information that can be extracted. It is very important to understand how to correlate, associate, map and bundle it with other sources of data to fabricate the required result. It also includes identification of the data source and the formats, thereby mapping the available information to given (known) result which may change after the discovery of different essentials and characteristics of data. There are various methods of Apriori Rule base consisting of association rule mining, clustering, classification, link mining and statistical learning, all of them playing remarkable roles in the development of data research applications. The use of the mining techniques and the mining equipment helps in boosting sectors like science and technology, education sector, business class and healthcare by providing simple access to assorted learning and designs of the market database for the better analysis of the market rising trends thus helping to opt for the corrective corner field to focus to focus upon [8].

B. Techniques: Data Mining

The Data mining techniques can be broadly classified as [9]:

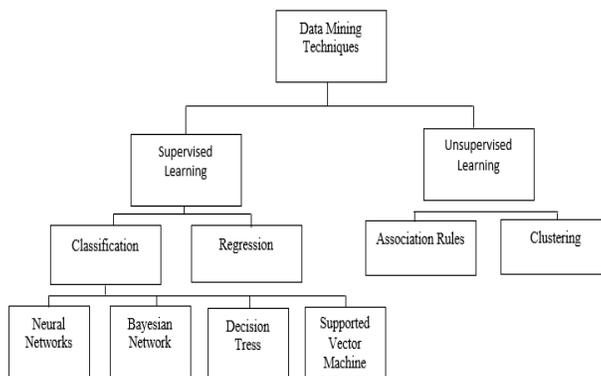


Fig.3 Data Mining Techniques

Supervised Learning: The majority of practical applications uses supervised form of learning. It is a type of learning where, there are two classes of variables. One is the input variable (x) and the other is the output variable (Y), and an algorithm is used to learn the mapping function that means from the input to the output given by: $Y = f(X)$. The aim is to approximate the mapping function such that whenever there is an input data available i.e. new input, then the

value can be easily predicted. Here, the data is labelled and used algorithms learn to predict the output from the available input data.

- a. Classification: It is a technique i.e. based on machine-learning process used to identify the items in the given data sets into one the predefined set of groups or classes. The classification analysis retrieves unique information about the meta-data, and the data [6]. It is helpful in classifying the distinct forms of data in different classes. In this type of analysis, the use of algorithms is necessary which decides how the new data should be classified.
 - Neural Networks (NN): In general, neural networks are called as Artificial Neural Networks that contains layers of interconnected nodes where each node for its input produces a non-linear function and the input to a node comes directly from the input source or from the other forms of nodes. Thus it helps in making simple calculations and recognizes different patterns.
 - Bayesian Network: It represents a graphical model that helps in encoding the probabilistic relationships among variables of interesting behavior. These networks when used in combination with statistical techniques, then the model represents several advantages for data modelling or designing. This model handles the situations where the data entries are not specified or missing. A Bayesian network can be used in learning of relationships (casual), and hence, identifies a problem domain that predicts the results of intervention. The model consists of semantics both casual and probabilistic that represents an ideal way of combining the data and prior knowledge. It also helps in avoiding the data overlapping problem.
 - Decision Trees: A decision tree denotes a structure which includes branches, a root node, and leaf nodes. Here, each of the internal node represents a test on an attribute, each of the branch provides the outcome of test performed, and the leaf node helps in holding a class label [5]. The decision tree constructs regression or classification model denoting a tree structure. It breaks the dataset into small subsets. On the other hand, an associated form of decision tree is developed incrementally. The final step results in a tree structure with decision and leaf nodes.
 - Supported Vector Machine: SVM primarily represents a classifier method that performs the tasks of classification by modelling or constructing the hyperplanes in a space i.e. multi-dimensional and further separates the cases of different class labels. SVM supports both classification and regression tasks and one can handle continuous variables easily. SVM can deal with the complexity of the model, and deals with problems based on real world such as hand-writing recognition, bio-sequence analysis, bio-informatics text, and image classification.
- b. Regression: The regression represents a method that analyse and models the relationships between variables and may contribute and relate to produce a specific outcome together. A simple case of regression is presented by a simple case of Single-Variable Linear Regression which is technique that linearly models a relationship between a single input (independent) and an output (dependent) variable [4]. There is another case of Multi-Variable Linear Regression which tells about the relationship between the independent multiple inputs i.e. feature variables and a dependent output variable.

A few lead points about the property of Linear Regression:

- Fast and easy type of modelling.
- Very easy to interpret and understand.
- The process of Linear Regression is sensitive to outliers.

Unsupervised learning: Here, only the input data (X) is available but there is no output for the process. The main motive of unsupervised form of learning is to design or model the distribution or the underlined structure in the data in order to gain more knowledge about the data. Here, the data is in unlabelled form and the algorithms inherits the structure from the input data.

- a. Association Rule Mining:

It represents the most important aspect of mining. It is helpful in detecting the patterns, correlations, casual structures or associations from the set of data that is found in databases such as transactional databases, relational databases, and other repositories available. The main aim is to search out the type of rule that predicts the specific item occurrence of another item. It includes a criterion to analyse the patterns that occur frequently based on two of its parameters. One is the 'support' and the other is the 'confidence' use to identify the relationship [6]. The support

denotes those type of data items that occur very frequently in the database where as the confidence part indicates the checks the truthiness of the statement again and again.

$$\text{Support, } S(X \cup Y) = \frac{\text{Support Sum of } XY}{\text{Complete Dataset of } d}$$

$$\text{Confidence, } C(X/Y) = \frac{\text{Support } XY}{\text{Support}}$$

b. Clustering:

The process of clustering is to make a group of abstract objects into the classes of objects that are similar in nature. A data cluster object can be treated as single (one) group. While performing the analysis, the first step is to make partition of dataset into groups that are based on data similarity and further assigns the labels to each group. The advantage of using clustering over classification process is that, it can be changed according based on requirement and helps in finding out single feature that clearly distinguish different forms of groups [3]. The basic requirements of clustering include the following:

- Scalability: There is a need of highly scalable algorithms that deal with large set of databases.
- Ability to deal with distinct attributes – The used algorithms must be capable such that they can be applied on any form of data such as categorical, binary data, and numerical data.
- Clusters discovery with shape of attribute: The algorithm of clustering must be capable to detect the arbitrary shape of the clusters. These are not bound to distance measures specifically as these tend to find only spherical form of cluster basically of small size.
- High dimensionality: The method clustering do not work on low dimensional features as these are to work on only high dimensional features.
- Ability to deal with data noise: Various forms of databases contain erroneous or noisy data. Some of the algorithms are very sensitive to such kind of data and it may lead to bad quality of clusters.
- Interpretability: The results of clustering should be easily interpretable, usable and comprehensible.

C. Prediction System: Heart Disease

The name heart diseases itself defines the term covering the inactive and disordered operation of the heart. These are well known as cardiovascular diseases (CVD's) forming the major cause of death globally. The death rate has increased annually from CVD's rather than any other cause. The year, 2015 describes the data of near about 17.7 million people who died from CVD's, covering 30% of the universal deaths. Out of these deaths, approximately 7.4% million died owing to heart strokes. A cardio attack strikes humans around every 43 seconds of time. When the flow of blood that specifically supplies oxygen to cardiac muscle pares in structure i.e. reduced then there are chances of experiencing a heart attack [12]. This generally happens because the blood flow to the arteries slows down thereby narrowing its structure leading to heart failure activities. The heart attack symptoms are intense and sudden in behavior. It mostly starts slowly causing discomfort and mild pain experienced in upper party of the body.

The major symptoms of heart stroke include the following:-

- Chest discomfort: The heart stroke generally arises from the centre of the chest lasting for a few minutes or more thereby causing acute pain in the chest area i.e. often repeated two or three times.
- Upper body discomfort: The symptoms include pain in the neck, arms (one or both) jaw, back and the stomach.
- Choking of breath: It includes severe chest problem and pain in the chest area.
- Other signs: These may cover the problems of nausea and burst forth in a cold sweat

The present scenario describes the most challenging problem faced by the medical research field. The technique of mining helps to predict the early stages of heart disease risks. Some of the major reasons of heart disorder embrace high cholesterol, pulse rate and blood pressure. The age of humans is non modifying risk factor in case of heart disease prediction. It also consists of some modifiable factors such as smoking and drinking. The heart forms the fundamental part of human body, it pumps new blood into the human body and the body might get paralyzed due to inadequate blood supply thereby affecting other parts of the body [11]. The over-stretching of blood vessels

increases the risk leading to high blood pressure problem which is guarded in terms of systolic and diastolic method of functioning. When the muscle gets contracted there is an acute coercion in the path of arteries indicating systolic behaviour whereas diastolic comprises of the coercion i.e. force in the path of arteries when the body is in resting state. The increased levels of fats or lipids that contains the structural component of living cells in the blood are the main reasons for heart attacks. The arteries contains lipids hence the blood flow in the arteries becomes very slow and is shaped in narrow state leading to the failure of heart and ultimately leading to death.

D. Risk Factors: Heart Disease Prediction

The uncertainty of heart stroke depends on large number of factors. The risk factors are habits or the conditions contributes to develop a disease assuredly which further increase the chances an existing disease to its hazardous form. Some of the factors are the once that are unchangeable in nature, such as family history and age of human [7]. The other factors are known as modifiable risk factors like cholesterol, smoking, uncontrolled hypertension, uncontrolled diabetes stress and anger, obesity, high blood pressure, lack of physical exercise, excess weight, social isolation and lack of support which can be changed when required. So, the advanced use of mining techniques of the data can help to overcome the unpredicted cardiac disease.

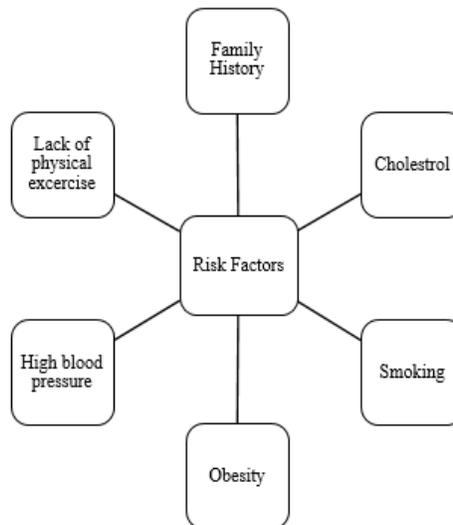


Fig.4 Risk Factors

The symptoms of a heart attack includes the following:

- Pain in the chest, discomfort, pressure, heaviness.
- Discomfort radiating to the throat, jaw, arm or back.
- Choking feeling, Fullness, and Indigestion
- Dizziness, Nausea, Sweating, or vomiting
- Extreme anxiety or weakness

II. RELATED WORK

Ming-Syan, et.al [1] conducted a research on Apriori Rule base knowledge and the information from a large set of databases. Here, a classification of data-mining was done with its available from and comparative study of such kind of techniques was presented. Fayyad, Usama, et.al [2] conducted a research on the process of Apriori Rule base and knowledge-based discovery in available databases. This article provides an overview of emerging research field which clarifies how the processes of Apriori Rule base and knowledge-based discovery in databases are related to each other, and to the other related fields like statistics, databases and machine learning. Bernard Buxton, et.al [3]

conducted a research forecasting the need for robust, accurate and fast algorithms for the analysis of the data. Palaniappan, et.al [4] conducted a study on Intelligent Heart Disease

Prediction System (IHDPS). This kind of research has developed a prototype named IHDPS using Apriori Rule base techniques like Neural Network, Naïve Bayes and Decision Trees. The results has shown that each of the technique has a specific strength of realizing the basic objectives of the Apriori Rule base methods. Shouman, Mai, et.al [5] conducted a research on Heart disease prediction methodology using several mining techniques. This research provides investigation that applied a range of techniques to various different forms of decision trees seeking better performance in heart ailment diagnosis criteria. The research proposed a model that outperforms Bagging algorithm and J4.8 Decision Tree in the diagnosis of heart disease patients. M. Sumender Roy, et.al [6] conducted a study based on Human Life Span Heart Disease that represents a major cause of mortality and morbidity. The outcome of the system reveals that the method of Decision Tree performs well and in most of the cases the classification done using Bayesian has similar accuracy as that of a decision tree. Bhatla, Nidhi, et.al [7] proposed a work on Heart disease prediction. The objective was to analyse the different techniques of Apriori Rule base for the prediction of heart disease. The experimental observation stated that neural networks with its 15 attributes has out-performed over all other mining techniques. It was also concluded from the analysis that the decision tree has also shown better form of accuracy with the use of genetic algorithm (GA) and feature- based subset selection. Smitha, T., and V. Sundaram [8] proposed an approach to analyse the implication of the dense data and low threshold value, how effective would be the rules of association. The data set which has been used represented a real time for data of some area and the dataset was applied in conjunction with the rules of association in order to predict the chance of heart disease hit by using Apriori algorithm. Kesavaraj, et.al [9] proposed a research on the process of Apriori Rule base The researchers have presented the basic forms of classification technique such as Bayesian networks, k-nearest neighbour classifier, decision tree induction etc. and the goal of such kind of study was to describe a review on comprehensive classification techniques in the Apriori Rule base analysis.

III. PROPOSED METHOD

T

he research methodology is the basic framework action plan adopted in carrying out the plan.

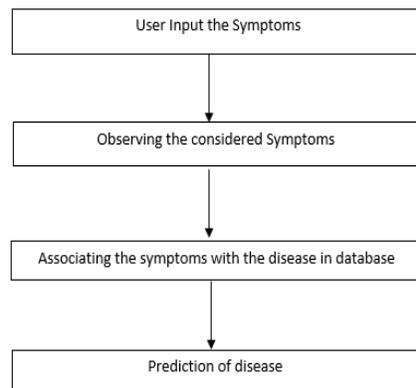


Fig.5 Flow Chart for Proposed System

To analyse the collected data, the basic tool used is R tool. The Apriori Rule base tool Weka 3.6.6 is used for experiment. Weka is a collection of machine learning algorithms for Apriori Rule base tasks. The algorithms can either be applied directly to a dataset or called from your own Java code.

Table 1:Accuracy

Techniques	Accuracy
Naive Bayes	86.53 %
Decision Tree	89%
ANN	85.53%

IV. RESULTS AND DISCUSSION

There are various responses that have been collected from the database of patients according to the formed questionnaire. The experimental analysis describes the heart disease analysis based in the demographic files based on the gender, age, disease status and it also shows the comparative result analysis based on positive prediction, sensitivity, accuracy, specificity. The response that is collected gets mined using the method of clustering and the analysis that is presented below:

- Gender

As listed in the Table.2 below, the majority of patients are males as compared to the females.

Table.2 Gender Demographic File

Demographic File	No. of patients (N=100)
Gender	Frequency
Male	82
Female	38

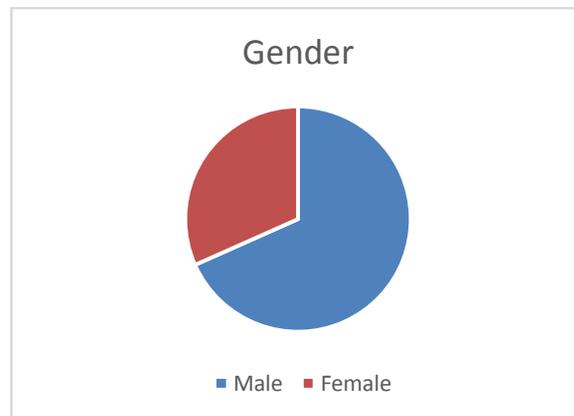


Fig.6 Gender

- Age

As listed in Table.3 highest propagation of the patients belongs to age group of 50-60 years, that is followed by <50 years, lowest proportion is of age 60<

Table.3 Age Demographic File

Demographic File	No. of patients(N=100)	
	Male	Female
Age(in years)		
<50	10	16
50-60	20	13
60<	52	9

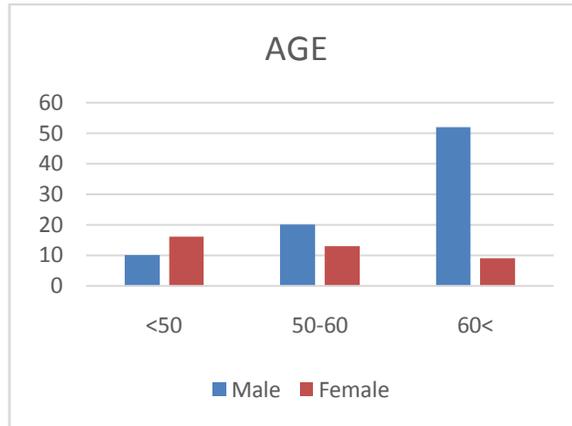


Fig.7 Age Profile

• **Disease Status**

The table.4 listed below represents the demographic file of the heart disease status for both the male and female in the form of positive, risk, and the negative factors

Table.4 Disease Status

Demographic File	No. of Patients (N=100)	
	Male	Female
Heart-Disease Status		
Positive	30	15
Risk	30	13
Negative	32	10

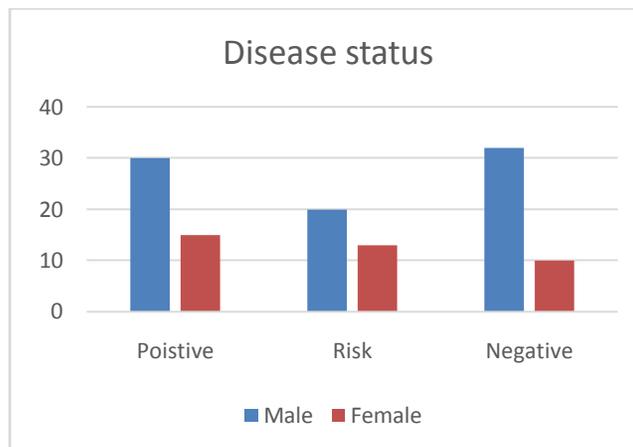


Fig.8 Disease Status

• **Positive Prediction**

In figure 9 comparison of positive prediction which show the true positive in proposed method it will increase of all classes because in proposed method rules are pruned by apriori classification approach.As listed below in table.5, positive prediction comparison has been done based on the proposed and the existing prediction methodology. Further, the prediction is divided into a comparative study of disease, risk and no type of disease analysis.

[Positive predictive value \(PPV\)](#), [Precision](#) =

$$\frac{\sum \text{True Positive}(TP)}{\sum \text{Predicted Condition positive}}$$

Table.5 Positive Prediction Comparison

Positive Prediction	Disease	Risk	No Disease
Proposed	66	70	99
Existing	56	69	98

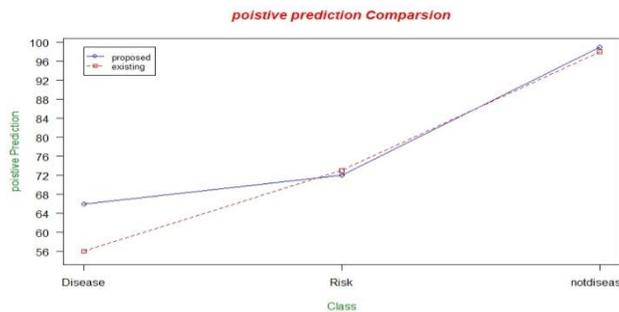


Fig.9 Positive prediction

• **Sensitivity**

In figure 10 comparison of sensitivity which show the false positive error in ration of true positive in proposed method it will increase of all classes because in proposed method rules are pruned by apriori classification approach. As listed below in table.6, the prediction based on sensitivity comparison has been done i.e. based on the proposed and the existing methodology. Where, P and FN stands for condition positive and False negative respectively.

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}$$

Table.6 Sensitivity Comparison

Sensitivity Prediction	Disease	Risk	No Disease
Proposed	92	99	93
Existing	88	90	89

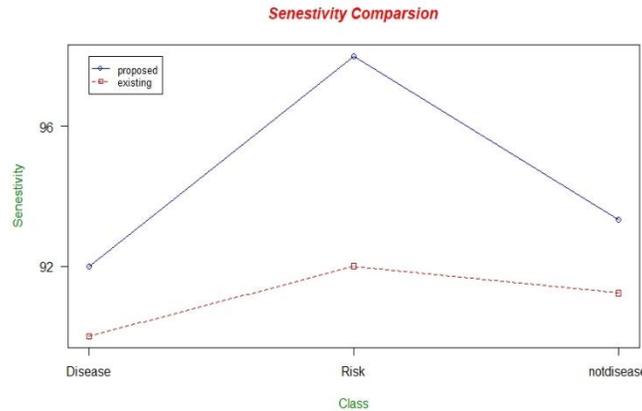


Fig.10 Sensitivity prediction

• **Specificity**

In figure 11 comparison of specificity which show the false negative error in ration of true positive in proposed method it will increase of all classes because in proposed method rules are pruned by apriori classification approach..As listed below in table.7, the prediction based on specificity comparison has been done i.e. based on the proposed and the existing methodology. Where, N, TN and FP stands for Condition Negative, True Negative and False Positive

$$TNR = \frac{TN}{N} = \frac{TN}{TN+FP}$$

Table.7 Specificity Comparison

Sensitivity Prediction	Disease	Risk	No Disease
Proposed	96	80	97
Existing	92	68	90

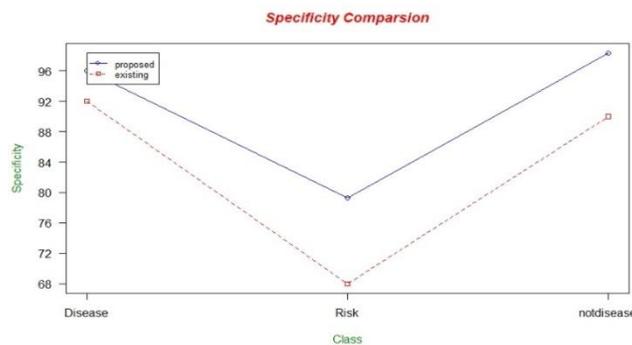


Fig.11 Specificity Comparison

• **Accuracy**

In figure 12 comparison of accuracy which show the true positive and true negative ratio with all instances in proposed method it will increase of all classes because in proposed method rules are pruned by apriori classification approach.As listed below in table.8, the prediction based on accuracy comparison has been done i.e. based on the proposed and the existing methodology.

$$\text{Accuracy} = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}$$

Table.8 Accuracy Comparison

Accuracy Prediction	Disease	Risk	Not Disease
Proposed	62	82	97
Existing	58	78	87

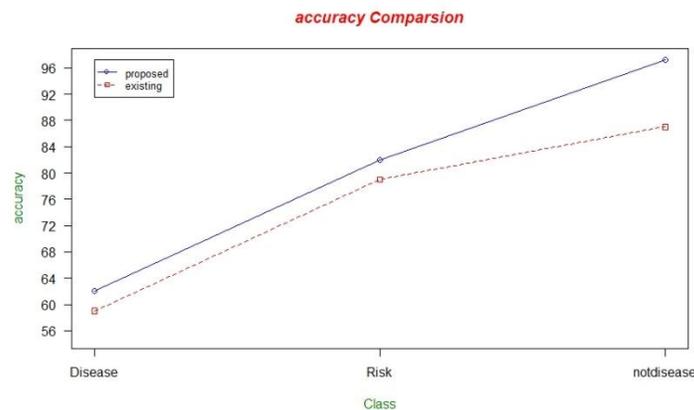


Fig.12 Accuracy Comparison

V. CONCLUSION

With the rowing development and increased quantity of data being generated by the researchers there is a big need for robust, accurate, and fast algorithms for the analysis of the data. Large improvements in the technology of databases, artificial intelligence, and computing performance and artificial intelligence have contributed towards the development of analysis of data. The main objective of Apriori Rule base is mainly to discover such patterns in data that leads to better and good understanding of the Apriori Rule base process further presenting useful predictions. Examples of Apriori Rule base applications includes character recognition (automated zip code) reading, detecting fraudulent transactions based on credit cards, and predicting compounded activity. So, this thesis work will propose a methodology based on heart disease prediction. The experimental analysis consists of demographic files (gender, age, disease status) and the comparative study based on different types of prediction i.e. positivity, sensitivity, specificity, and accuracy has been done to improve the analysis. Finally, the result obtained are compared which shows that proposed methods outperforms that the results based on the existing methods.

REFERENCES

1. Kumar, "Early Heart Disease Prediction Using Data Mining Techniques," department of Computer Science & Information Technology014 DOI : 10.5121/csit.2014.4807 pp. 53–59,
2. Data Mining - Classification & Prediction. 2016. Available:https://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm.
3. Stag log, "Testing the Diagnosis", Available at: <https://www.thehindu.com/opinion/op-ed/testing-the-diagnosis/article22708173.ece>
4. B. Bahrami, and M. H. Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques," *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, ISSN: 3159-0040, Vol. 2, Issue 2, February 2015.
5. I. H. Witten and E. Frank, "Data Mining Practical Machine Learning Tools and Techniques," Morgan Kaufman Publishers, 2005.

6. A. K. Pandey, P. Pandey, K. L. Jaiswal, and A. K. Sen, "Data Mining Clustering Techniques in the Prediction of Heart Disease using Attribute Selection Method," *International Journal of Science, Engineering and Technology Research (IJSETR)*, ISSN: 2277798, Vol 2, Issue10, October 2013.
7. Isra 'a Ahmed Zriqat, Ahmad Mousa Altamimi, Mohammad Azzeh, "A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods," *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 14, No. 12, December 2016
8. Ahmad, P., S. Qamar, and S.Q.A. Rizvi, *Techniques of data mining in healthcare: A review. International Journal of Computer Applications*, 2015. 120(15).
9. *A Comparative Study of Classification Techniques in Data Mining Algorithms. Oriental Journal of Computer Science & Technology*. 8: p. 13-19.
10. Mirpouya Mirmozaffari, Alireza Alinezhad, and Azadeh Gilanpour, "Data Mining Apriori Algorithm for Heart Disease Prediction," *nt'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE)* Vol. 4, Issue 1 (2017) ISSN 2349-1469 EISSN 2349-1477
11. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Ankromita Mukherjee and Asmita Mukherjee, "Herart Disese Diagnosis And Prediction Using Mahine Learning And Data Mining Techniques: A Review," *Advances in Computational Sciences and Technology* ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 2137-2159
12. *Data Mining - Classification & Prediction*. 2016. Available:https://www.tutorialspoint.com/data_mining/dmclassification_prediction.htm.
13. Aswathy Wilson, Gloria Wilson, and Likhya Joy K, "Heart Disease Prediction Using the Data mining Techniques," *International Journal of Computer Science Trends and Technology (IJCST) – Volume 2 Issue 1, Jan-Feb 2014*
14. Asha Rajkumar, G.Sophia Reena, "Diagnosis Of Heart Disease Using Datamining Algorithm," *Global Journal of Computer Science and Technology* 38 Vol. 10 Issue 10 Ver. 1.0 September 2010.
15. Stag log, "Heart Disease Data Set", Available at: <https://www.nhlbi.nih.gov/health/educational/hearttruth/lower-risk/risk-factors.htm>
16. Sellappan Palaniappan Rafiah Awang, *Intelligent Heart Disease Prediction System Using Data Mining Techniques, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008*
17. Kathleen H. Miao1 , Julia H. Miao1 , and George J. Miao2, "Diagnosing Coronary Heart Disease Using Ensemble Machine Learning (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 10, 2016
18. K.Srinivas, B.Kavihta Rani , A.Govrdhan , "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks," (IJCE) *International Journal on Computer Science and Engineering* Vol. 02, No. 02, 2010, 250-255.
19. Latha Parthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm,"
20. *Internatinal Journal of Biological, Biomedical and Medical Sciences*, Vol. 3, No. 3, 2008.
21. Quinlan J. *Induction of decision trees. Mach Learn* 1986; 1:81—106.
22. Mai Shouman, Tim Turner and Rob Stocker, "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.
23. Gupta, G., Aggarwal, H and Rani, R. (2015), "Mining the Customers Data for making Segments based on RFM Analysis in Building Successful and Profitable CRM", *Ciência e Técnica Vitivinícola Journal*, Vol. 30, No. 3, pp. 261-270.
24. Gupta, G. & Aggarwal, H. (2012), "Improving Customer Relationship Management Using Data Mining", *International Journal of Machine Learning and Computing (IJMLC)*, Vol. 2, No. 6, pp. 874-877.
25. Gupta, G. & Aggarwal, H. (2016), "Analyzing Customer Responses to Migrate Strategies in Making Retailing and CRM Effective", *Int. J. Indian Culture and Business Management (IJICBM)*. Vol. 12, No. 1, pp. 92-127.
26. Gupta, G., Aggarwal, H. & Rani, R. (2016), "Segmentation of Retail Customers based on Cluster Analysis in Building Successful CRM", *Int. J. of Business Information Systems (IJBIS)*. Vol. 23, No. 2, pp. 212-228.

27. Gupta, G. & Kahlon, J.S. (2016), “PREDICTING THE CAUSE OF ABSENTEEISM AMONG PUBLIC VERSUS PRIVATE COLLEGE STUDENTS USING DATA MINING”, *International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM)*. Vol. 4, No. 4, pp. 1-5.
28. Kaur, S. & Gupta, G. (2016), “DATA MINING APPROACH TO CRM IN BANKING SECTOR”, *International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM)*. Vol. 4, No. 2, pp. 81-89.
29. Kaur, N. & Gupta, G. (2016), “EFFECTIVENESS OF CRM IN RETAIL SECTOR USING DATA MINING”, *International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM)*. Vol. 4, No. 2, pp. 71-80.
30. Kaur, K. & Gupta, G. (2015), “PREDICTING THE USE OF INTERNET AMONG TEACHERS AND STUDENTS USING DATA MINING”, *International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM)*. Vol. 3, No. 3, pp. 97-106.
31. Kaur, N.K. & Gupta, G. (2015), “PREDICTING THE VARIOUS RISK FACTORS OF LEPROSY USING DATA MINING TECHNIQUES”, *International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM)*. Vol. 3, No. 3, pp. 79-87.
32. Kaur, H. & Gupta, G. (2014), “Optimizing Cart Algorithm by Enhancing Attributes”, *International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM)*. Vol. 2, No. 3, pp. 34-38.
33. Kaur, J. & Gupta, G. (2014), “Hybrid of K-Means and Hierarchal Algorithms to Optimize Clustering”, *International Journal for Multi Disciplinary Engineering and Business Management (IJMDEBM)*. Vol. 2, No. 3, pp. 39-44.