Observation of Data Mining in Healthcare

[1]MANDAKINI PRIYADRSHINI BEHERA, *Gandhi Institute of Excellent Technocrats, Bhubaneswar, India*
[2]SASWATI JENA, *Kalam Institute of Technology,Berhampur, Ganjam, Odisha, India*

ABSTRACT: Data mining is a field of study that entails extracting usable information or knowledge from previously collected data. Data mining is done using a variety of approaches. Data mining can be applied to a variety of industries, including healthcare. Heart or cardiovascular illnesses are a hot topic in the worldwide healthcare business. Many patients perished as a result of a lack of knowledge. Because the healthcare business generates so much data, we can utilize data mining to uncover hidden patterns and fascinating information that can aid in making more informed decisions. Data mining in healthcare is a critical and difficult task that must be completed correctly. It aims to address real-world health issues such as disease diagnosis and treatment. This research also aims to uncover interesting patterns in data from heart patients. With two separate cases, three algorithms are applied. Decision Tree, Neural Network, and Nave are the implemented algorithms.

1 INTRODUCTION

Data mining, also known as Knowledge Discovery in Databases (KDD) [1], is one of the most popular areas of research, involving the discovery of interesting patterns and useful data. Various Data Mining tools and approaches are used to predict data behavior and trends, allowing specialists to make more active and accurate decisions based on their knowledge. Data mining has been successfully employed in a variety of industries, including marketing, banking, and enterprises; but, it might also be applied in healthcare.

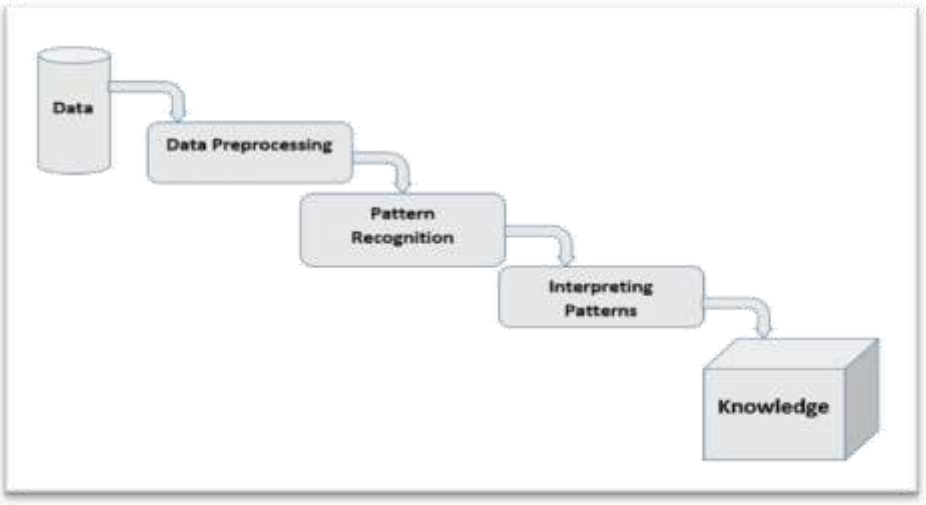
Heart or cardiovascular disorders are currently a very popular topic in the worldwide healthcare business. The most recent data [2] was published by the World Health Organization (WHO) in April 2011. Coronary Heart Disease Deaths Have Reached Record Highs, according to the data. Health According to that, coronary heart disease fatalities accounted for 15.36% of all deaths in Pakistan, and the World Health Organization estimates that by 2030, more than 23 million people will die from heart disease yearly [3].care industry produces a huge amount of data about heart diseases but all is waste and nothing as it does not helpsin effective and efficient decision making. Doctors and health care experts each have their own experiences that they use to anticipate a patient's specific heart illness, which can sometimes lead to incorrect results. So, utilizing the patient's data, data mining techniques must be deployed and analyzed to uncover hidden patterns that can help in a variety of ways. Each data mining approach serves a different function based on the requirement and usage. Although data mining in healthcare is crucial, it is a difficult task that must be completed correctly. It aims to address real-world health issues such as disease diagnosis and treatment.

There will be many benefits of Data Mining in Healthcare, such as grouping patients with similar diseases or health issues so that effective treatments can be provided, checking or providing availability of medical solutions to patients at lower costs, safe healthcare treatment, reducing the time for medical treatment, detecting disease causes and identifying medical treatment methods, and efficient data mining. It also aids healthcare providers and specialists in developing effective healthcare strategies.

We will utilize .Weka data mining tools to analyze an online dataset of cardiac patients for our research. We test the performance of three distinct methods, including Decision Tree, Neural Network, and Nave Bayes, with and without attribute selection.

A SUMMARY OF DATA MINING

Because of the vast amount of data available and the need to translate it into valuable knowledge, Techniques like data mining can be beneficial. Data mining has been increasingly important in practically every industry, including health care, in recent years. The scenario is characterized as "data rich but information poor" because of the volume of data and the necessity for effective analytical tools for that data. Figure 1 depicts a knowledge flow diagram.



1 APPLICATION OF DATA MINING IN THE HEALTH SECTOR

Without a doubt, the health sector [4] has a greater need for data mining nowadays, and data mining technologies are gaining prominence in that industry. Successful data mining applications in the health industry have been recorded, and it has enormous potential. For example, data mining tools can compare and contrast causes and symptoms to determine the effectiveness of medical therapies, as well as uncover successful standardized treatments for specific diseases. These can also aid in health-care management by identifying and tracking disease states and risks to patients. Another key task in the health sector is to increase quality at a cheap cost [5]. Data mining is used by healthcare insurers to lessen the need for medical insurance. Data mining is used by healthcare insurers to reduce medical insurance fraud and misuse. Because diagnostic and laboratory procedures are expensive and painful for patients, data mining is becoming more popular in the medical field for diagnosis and decision support.

One of the limitations of data mining in the health care sector is the integration and collection of data; raw data for data mining can be found in a variety of places and systems; the solution is to collect all of the data in a data warehouse. The third factor is data quality; efficient results require quality data, and missing, inconsistent, or non-standardized data in various formats can lead to a lack of quality. Domain expertise, as well as knowledge of data mining tools and methodologies, is essential. Finally, healthcare organizations must devote significant resources, time, effort, and money to mining, and mining teams must collaborate. Data mining necessitates meticulous planning and technology setup.

1 HEART DISEASE OVERVIEW

The efficient working heart is completely dependent on the human life without any breaks or pauses. Heart disease, often known as cardio disease, is an illness that affects the heart and its blood circulation system. It refers to a wide range of diseases and disorders that affect the heart and, in certain cases, the blood vessels. These are caused by heart and pumping system problems. This can lead to sickness, disability, or even death.

FACTORS THAT CAUSE HEART DISEASE AND HOW TO PREVENT THEM

A lot of variables contribute to the development of heart disease. Heart issues and disorders are caused by these variables. Family history, hypertension, blood pressure, cholesterol, smoking or tobacco use, poor or unhealthy diet, physical inactivity, and so on are some of these factors. Many heart diseases can be avoided in their early stages by simply preventing or controlling them. Preventative strategies include regular exercise, a healthy and well-balanced diet, quitting smoking, maintaining a normal healthy weight, and so on. Diabetes, cholesterol, hypertension, and other risk factors can be controlled or prevented with regular medical care and a change in lifestyle.
Hospitalization was required for critical cardiac disease such as heart attack, heart failure, or stroke. Heart attacks, heart failure, and stroke all necessitate hospitalization, and the treatment for these disorders includes increasing the amount of oxygen delivered to the heart tissues. It also comprises vital sign monitoring and advanced life support procedures.

1 .**WORKS IN CONNECTION**

Using data mining techniques, a lot of work has been done on cardiac problems. The following are a few of them.
Enhanced Heart Disease Prediction with Feature Subset Selection Using Genetic Algorithm [6] was proposed in 2010. The Genetic Algorithm is a natural selection and genetics-inspired optimization tool. A heart disease system was created to accurately predict the existence of heart disease using fewer variables. They improved heart disease prediction with their suggested method by applying feature-based subset selection instead of the 13 attributes that were previously employed. There are three different classification systems. For the diagnosis of patients with heart disease,
There are three different classification systems. Patients with heart disease were diagnosed using decision trees, classification with clustering, and Naive Bayes, with trials using the Weka data mining tool. In comparison to other Nave Bayes and Classification by Clustering, these trials reveal that the Decision tree has the highest accuracy and construction time.

HDPS: Heart Disease Prediction System [7] was introduced in 2011, and it uses only one data mining algorithm, the artificial neural network (ANN), to classify heart disease based on 13 different attributes. This system uses a data set of 303 instances from the UCI machine learning repository.
Data mining neural network technique was utilized for heart disease prediction in 2012 [8]. This demonstrates near-perfect accuracy. The system uses a Multilayer Perception Neural Network (MLPNN) with Back Propagation Algorithm (BP), which is one of the most important models in neural networks. It is made up of many levels that are connected by little circles known as nodes. Back propagation is a widely used algorithm.

The back propagation algorithm estimates the difference between actual and expected values from output nodes to the preceding node layer. The experiments are conducted using the WEKA data mining programmed, and the data set contains 573 records divided into two parts: training and testing. To improve the accuracy of the forecast, a total of 15 attributes were used.

Weighted Associative Classifiers (WAC) [9] were established in 2013 as a clever and effective heart disease prediction method. In WAC, different weights are applied to properties based on their capacity to predict. The system is built on the JAVA platform, with data from the UCI repository available online. There are 303 records in the data set, with 14 different properties. Experiments have shown that WAC is approximately 80% accurate.
Various data mining algorithms were applied to forecast heart disorders in 2013 [10]. The Weka tool was used to undertake several tests. Ten cross validations with and without bagging were used to compare the findings of these tests. Bagging stands for Bootstrap aggregation, which is used to improve classification accuracy. There were three techniques utilized, and In this study, three strategies were used and compared: Nave Bayes, J48 Decision Tree, and Bagging. The data set came from the Hungarian Institute of Cardiology and included 76 raw variables, however only 11 were chosen for testing. Bagging has the highest accuracy (85.03%), whereas J48 decision tree and nave bayes have 84.35 percent and 82.31 percent accuracy, respectively.

1 MATERIALS AND METHODS

DATA SET FOR EXPERIMENTS

We took online available heart disease datasets from UCI (University of California, Irvine C.A) Machine Learning Repository [12]. These dataset have the same instance format and attributes. These datasets have 76 raw attributes including predicted attribute but only 14 of them are actually most important. Cleveland Clinic Foundation data set contain 303 patients record and Hungarian Institute of Cardiology data set contains 294 patient records. We integrate both of these datasets. So it means that there are total 597 instances in our dataset. The attributes of the dataset and their description is given in Table 1.

Table 1. Attributes and their Descriptions

Serial #	Attributes	Description	Values	Type
1	Age	Patient's age in years		Numeric
2	Sex	Sex of Patient	Male, Female	Nominal
3	ch_pain	chest pain	Value 1: typical angina Value 2: atypical angina Value 3: non-anginal pain Value 4: asymptomatic	Nominal
4	r_B_Pressure	resting blood pressure (in mm Hg on admission to the hospital)		Numeric
5	Chol	serum cholesterol in mg/dl		Numeric
6	f_B_sugar	fasting blood sugar > 120 mg/dl	1 = true 0 = false	Nominal
7	r_ECG_results	Resting electrocardiographic results	0: normal 1:having ST-T wave abnormality (T wave inversions and/or STelevation or depression of > 0.05 mV) 2: showing probable or definite left ventricularhypertrophyby Estes' criteria	Nominal
8	maxi_heart_rate	maximum heart rate achieved		Numeric
9	Exercise	exercise induced angina	1 : yes 0 : no	Nominal
10	Oldpeak	ST depression induced by exercise relative to rest		Numeric
11	Slope	the slope of the peak exercise ST segment	1: upsloping 2: flat 3: down sloping	Nominal
12	no_of_majorvessels	number of major vessels (0-3) colored by fluoroscopy		Numeric
13	Defectype	defect type	3 = normal 6 = fixed defect 7 = reversible defect	Nominal
14	Class	diagnosis of heart disease	no_heart_disease have_heart_disease	Nominal

FILLING THE MISSING VALUES

The dataset for our experiments had some missing records, those missing records were find and replaced with appropriate value using ReplaceMissingValues filter from Weka tool. This filter scans all the records and replaces the missing values by mean mode method.

ATTRIBUTE SELECTION DATA REDUCTION TECHNIQUE

Attribute selection or Feature selection is a data reduction technique that is applied to the dataset. This technique reduces the size of the dataset by removing the irrelevant or redundant attributes. Other than this attribute selection method has other benefits such as the reduce number of attributes helps to make patterns easier and understandable. The Selected Attributes in Weka using BestFirst search method is shown in Figure 1.

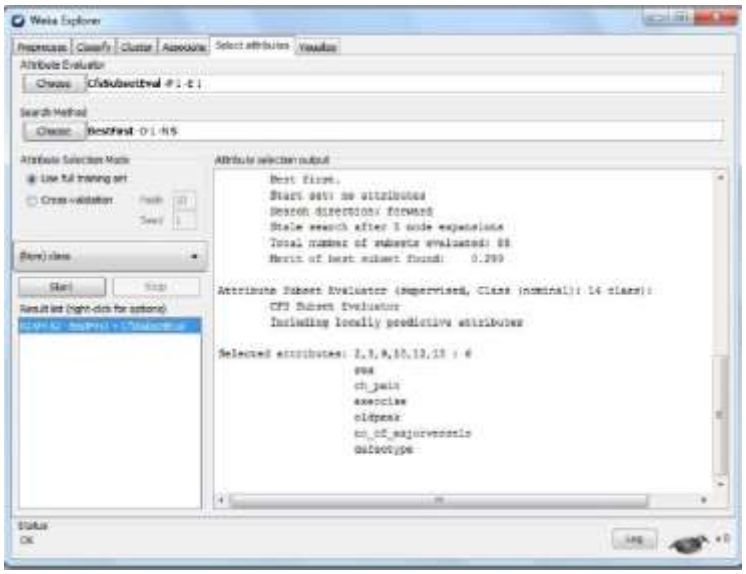


Fig. 2. Attribute Selection in Weka

For our experiments we will select CfsSubsetEval attribute filter. This filter evaluates the worth of subset of attributes in the given dataset by considering the individual predictive ability of each feature and with this filter we will use the Best First searching method. In this method, after each of the iteration the best of the original remaining attributes are added to the dataset. This method selects 7 attributes from the total of 14 attributes. These selected attributes are sex, ch_pain, exercise, oldpeak, no_of_majorvessols and defectype.

WEKA MACHINE LEARNING SOFTWARE

WEKA (Waikato Environment for Knowledge Analysis) developed at the University of Waikato, New Zealand which is a popular suite of machine learning software. Weka [13] is freely available software under the GNU General Public License that is written in java.

Weka workbench contains a collection of visualization tools and state of the art machine learning algorithms for data analysis and modeling. Different data mining tasks are supported by Weka these tasks are data preprocessing, clustering, classification, regression and visualization. Weka support data set in ARFF (Attribute-Relation File Format) format. ARFF file has two section first section contains Header information and second section contains Data information.

ALGORITHMS USED FOR EXPERIMENTS

DECISION TREE

Decision Tree is a flowchart like tree structure which contains leafs, nodes and branches [14]. Decision Tree has become popular in knowledge discovery because the construction of decision tree classifier does not require any domain knowledge. Successful decision tree model depends upon the data but in general it has good accuracy. The sample decision tree is shownin Figure 3.

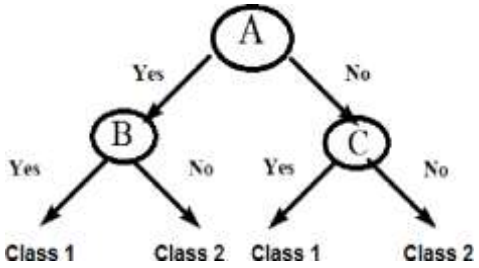


Fig. 3. Sample Decision Tree

J-48 is a type of decision tree algorithm that generate pruned and un-pruned decision tree for classification of data. The un-pruned trees are simply the larger in size where as pruning is a feature of J-48 decision tree that used to correctly classifying the subsets in the dataset. When a decision tree is built many branches shows anomalies in data due to outliers and tree pruning address this problem in data. This method removes the least reliable branches. It follows the facts in which each attribute of data can be used by splitting the data into subsets [15]. It uses the concept of information entropy. To makethe decision the attribute with highest information gain is used and information gain is basically the difference in entropy. This algorithm can handle both continuous and continual attributes.

ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks are the human neurons type network structure which consists of number of nodes that are connected through directional links where each node represents a processing unit and the links between them specify the casual relation between them. This classification technique is becoming powerful tool in data mining and may be used for different purposes in descriptive and predictive data mining. The sample artificial neural network is shown in Figure 4.

.

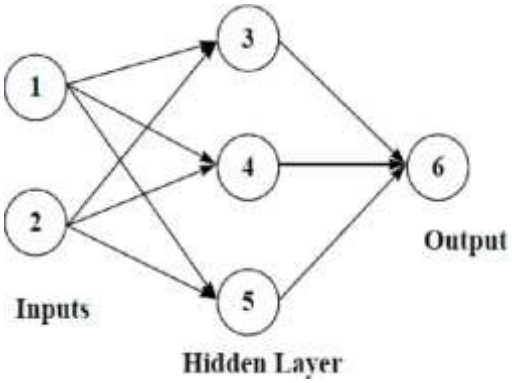


Fig. 4. Sample Artificial Neural Network

Many studies [16] shows that Artificial Neural Networks used in clinical decision making and helps the doctors to analyze and make sense if complex clinical data and medical applications. A Neural Network start with an input layer where eachnode is corresponds to a predictor variable [17]. The nodes of input layer are connected to the nodes of hidden layers and the nodes of hidden layer may be connects to each other or to an output layer. Multilayer Perceptron is a feed forward neural network with one input and output layer with several possible hidden layers that are totally interconnected.

BAYESIAN CLASSIFIERS

Bayesian classifiers are the classifiers that based on Bayes theorem [18]. They can predict about class probabilities such aswhich record is belong to a particular class and shows high speed and accuracy when applied to large databases.

Naïve Bayes algorithm is easy to construct and easy to interpret. It generally performs well as compare to other methods. Naïve Bayes classifier is based on conditional probability rule. It uses all the attributes in given dataset and analysis them individually.

2 EXPERIMENTAL SETUP

For our experiments we use Weka machine learning tool and we will select EXPLORER tab. Different experiments are conducted on full dataset. This dataset contains 597 unique attributes. We do experiments under two different scenarios, the first scenario is that in which all the attributes are used for experiments and the second scenario is that in which some most suitable attributes are selected using attribute selection method. The Visualization of the Dataset is shown in Figure 5.

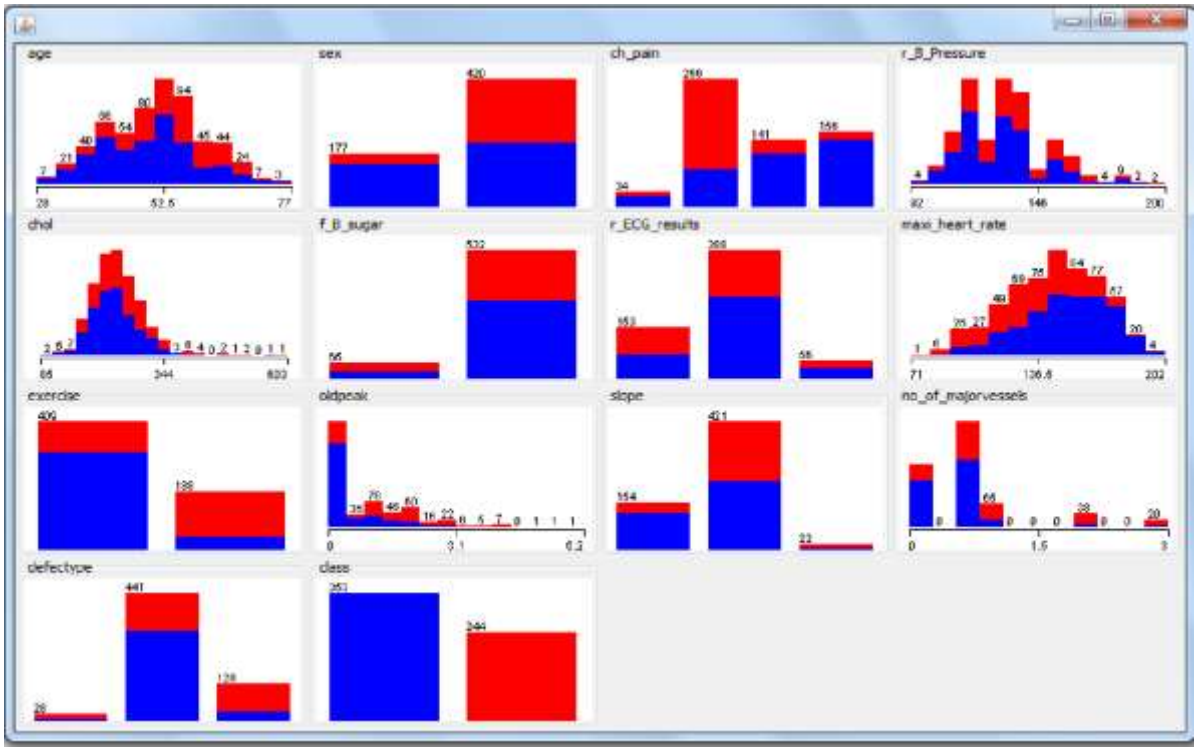


Fig. 5. Visualization of the Dataset

Mainly three algorithms are used in our bench mark data for experiment on both of the scenarios that are Decision Tree, Neural Network and Naïve Bayes. For sampling the training and testing dataset we use 10-fold cross validation. The performance and the accuracy of each experiment is evaluated by standard metrics such as TP rate, TN rate, precision, recall and F-measure which are calculated by Confusion Matrix which is known as predictive classification table. All these measures will be used to compare the performance of these selected and implemented algorithms.

3 DISCUSSION

The focus of our study was on using data mining in healthcare for heart diseases. For this purpose we do some experiments on our data about heart disease patients by applying different data mining algorithms. By implementingdifferent classification algorithms we try to find out which one from the following algorithms is best in predicting about heart diseases.

There are four experiments were performed and these experiment are designed for the different purposes, thesepurposes are to investigate the difference between pruned and un-pruned decision tree, the affects of attribute selection method on implemented algorithms and to compare the results of J-48 Decision Tree, Neural Networks and Naïve Bayes.

COMPARISON OF IMPLEMENTED ALGORITHMS

After the experiments the next step is to compare algorithms used in these experiments and select the best one, as we already know every experiment conducted in two different scenarios, in first scenario all the attributes were used and in second scenario only the selected attributes were used. To compare these experiments different performance measures are taken into consideration like accuracy, TN and TP rate, ROC curve and the time taken to built a model. The summery of implemented algorithms is given in Table 2.

Fig. 6. Summary of Implemented Algorithms Performance

Algorithm	Accuracy (%)	TP rate	FP rate	Precision	F-measure	ROC Curve	Time (sec)
J48 decision tree (pruned) with all attributes	78.057	0.781	0.253	0.779	0.778	0.786	0.11
J48 decision tree (pruned) with selected attributes	78.727	0.787	0.246	0.786	0.785	0.778	0.04
J48 decision tree (un-pruned) with all attributes	77.219	0.772	0.257	0.770	0.770	0.788	0.03
J48 decision tree (un-pruned) with selected attributes	79.062	0.791	0.233	0.789	0.789	0.821	0.01
Multilayer perceptron with all attributes	79.89	0.799	0.214	0.799	0.799	0.867	8.24
Multilayer perceptron with selected attributes	80.402	0.804	0.218	0.803	0.803	0.870	3.73
Naïve Bayes with all attributes	82.914	0.829	0.194	0.828	0.828	0.898	0.03
Naïve Bayes with selected attributes	82.077	0.821	0.197	0.820	0.820	0.898	0.01

In general the results of all the implemented algorithms with selected attributes are much better than algorithms with all attributes except Naïve Bayes. The table above presents the great accuracy of implemented algorithms on the given heart disease dataset, the lowest accuracy is 77.219% and the highest accuracy is 82.914%. The Naïve Bayes classifier algorithm with all attributes shows the highest accuracy i.e. 82.914 % and Navie Bayes with selected attributes is nearest to it with
82.077 % accuracy. On the other hand, J-48 decision tree (un-pruned) with all attributes score the lowest accuracy i.e. 77.219
%.

If we look at the time (sec) taken to built a model for selected algorithm we will find that the Neural Networks takes the longest time while Decision Tree and Naïve Bayes takes shortest time in building the model.

Other performances measures like TP rate and FP rate that are also used compare the results also achieve remarkable performance and are shows very close difference. The TP rate and the FP rate were, (0.781, 0.253) for J-48 (pruned) with all attributes, (0.787, 0.246) for J-48 (pruned) with selected attributes, (0.772, 0.257) for J-48 (un-pruned) with all attributes, (0.791, 0.233) for J-48 (un-pruned) with selected attributes, (0.799, 0.214) for multilayer perceptron with all attributes,(0.804, 0.218)for multilayer perceptron with selected attributes, (0.829, 0.194) for naïve bayes with all attributes, and (0.821, 0.197) for naïve bayes with selected attributes respectively.

This shows that the Naïve Bayes with all attributes score the highest TP rate i.e. 0.829 while J-48 (un-pruned) Decision Tree with all attributes score the lowest TP rate. We compare the entire TP rate and FP rate scored by all the algorithms we found that all of these algorithms were better in predicting positive cases as TP rate in them is always greater than FP rate.

Precision score and F-measure of all the algorithms were quite balance, the highest precision scored are 0.828 and 0.820 by Naïve Bayes with all attributes and Naïve bayes with selected attributes respectively while Multilayer Preceptron with selected attributes was closet by 0.803 precision score. If we look at ROC curves we found that the Multilayer Precptron and Naïve Bayes are relatively close and accurate algorithms. ROC curve value for Naïve Bayes classifier was 0.898 which is nearest the "Perfect Classification Point" i.e. 1.

Based on above results and comparisons we found that the Naïve Bayes performs the highest Accuracy, TP rate, Precision, F-measure and ROC curve value. Naïve Bayes also score the fastest execution time as compare to other algorithms. Multilayer Perceptron take the longest time to build a model.

We know heart disease is a fatal disease which causes the death of millions people annually so there is need to keep the number of True Positives high and number of False Positives low. Early diagnosis is the key factor for a successful treatment of a disease, therefore classification algorithms are expected to perform well and emphasis is given to select the algorithms having high TP rate. Accuracy is also worth considering that is accurately identifying heart patients as much as possible.

Based on these results and comparisons we will select Naïve Bayes classifier with all algorithms is selected algorithms to predict the heart diseases for our study. The experiments shown that Naïve Bayes algorithm outclass the Decision Tree and Neural Networks in the domain of predicting of heart diseases.

EFFECT OF ATTRIBUTE SELECTION METHOD

Our all the experiments are consists of two scenarios, one scenario with all 14 attributes and the other scenario with 8 selected attributes. Attribute selection method helps in increase of classification accuracy and decrease the time and complexity by ignoring irrelevant attributes from given dataset. The results of experiments shows that the accuracy of the implemented algorithms increased and the execution time was decreased by using attribute selection method.

SELECTED ALGORITHM BASED ON PERFORMANCE

Different objectives were defined at the start of our research. Those were evaluated against our selected algorithm. Results of the experiments show that the selected algorithms had achieved the highest score which suggest that this algorithm better in heart disease diagnosis.

4 FUTURE RECOMMENDATION

Our research indicates that data mining can be used and applied in healthcare industry to predict about heart diseases and implemented algorithms are worth considering. Further research should be conducted to increase the classification accuracy by using different classification algorithms such as Bagging, Support Vector Machine or Decision Table etc.

In our experiments we do not change the default parameters of algorithms, in future this study can be enhanced and expand by changing the parameters for experiments. The patterns and relationships found in heart disease dataset can be used to design a complete Knowledge Based system.

In future, more work can be done by using more data set related to heart diseases and by using different data reduction methods to improve the classification. For better accuracy and prediction of heart diseases the datasets that will be used must be quality oriented and free from outliers, inconsistencies and missing values.

5 CONCLUSION

Our study was focused on the use of data mining techniques in healthcare specifically in Heart Diseases. Heart disease is a fatal disease which may cause life threatening complications such as death. We use online available heart patient's data from UCI repository. There were 597 unique instances in our data set. The classification which is a data mining technique was implemented with following algorithms, Decision Tree, Neural Networks and Naïve Bayes. Some important point were considered to choose suitable tool for mining, on the basis of them Weka machine learning software were used for experiments. To evaluate the performance of the algorithms different performance metrics were considered that are accuracy, precision, F-measure, ROC curve value, TP rate and FP rate.

Four experiments were conducted in two different scenarios, in first scenario all attributes were used and in second scenario selected attributes were used and data set was in ARFF format that is supported by Weka. The experiments show that Naïve Bayes classification algorithms have the highest accuracy among all that is 82.914%. This study shows that the data mining can be used to predict about heart disease efficiently and effectively. The results or the outcomes of our thesis may be used as assistant tool to help in making more consistent diagnosis of heart diseases.

REFERENCES

[1] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques 2nd Edition" The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, March 2006.
[2] WORLD HEALTH RANKING 2014 [online] available: http://www.worldlifeexpectancy.com/cause-of-death/coronary-heart-disease/by-country/ (May 7, 2014).
[3] WORLD HEALTH ORGANIZATION 2014 [online] available: http://www.who.int/cardiovascular_diseases/en/ (May 7, 2014).
[4] Hian Chye Koh and Gerald Tan, "Data Mining Applications in Healthcare" Journal of Healthcare Information Management — Vol. 19, No. 2, 2005.

[5] Greg Rogers and Ellen Joyner, "MINING YOUR DATA FOR HEALTH CARE QUALITY IMPROVEMENT", SAS Institute, Inc., Cary, NC, January 1, 2009.
[6] M. Anbarasi, E. Anupriya and N.Ch.S.N.Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm" International Journal of Engineering Science and Technology Vol. 2(10), 2010.
[7] AH Chen, SY Huang, PS Hong, CH Cheng, EJ Lin, "HDPS: Heart Disease Prediction System", IEEE-Computing in Cardiology, September 2011.
[8] Chaitrali S. Dangare and Sulabha S. Apte, "A DATA MINING APPROACH FOR PREDICTION OF HEARTDISEASE USING NEURAL NETWORKS" International Journal Of Computer Engineering & Technology, Volume 3, Issue 3, 2012.
[9] Jyoti Soni, Uzma Ansari and Dipesh Sharma, "Intelligent and Effective Heart Disease Prediction System using WeightedAssociative Classifiers" International Journal on Computer Science and Engineering, Vol. 3 No. 6 June 2011.
[10] Vikas Chaurasia and Saurabh Pal, "Data Mining Approach to Detect Heart Dieses", International Journal of Advanced Computer Science and Information Technology Vol. 2, No. 4, 2013.
[11] S.Saravanakumar and S.Rinesh, "Effective Heart Disease Prediction usingFrequent Feature Selection Method" International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, March 2014.
[12] UCI (University of California, Irvine C.A) Machine Learning Repository, 2014 [online] available: http://repository.seasr.org/Datasets/UCI/arff/ (June 1, 2014).
[13] Remco R. Bouckaert, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann, Alex Seewald and David Scuse "WEKA Manual for Version 3-6-8" THE UNIVERSITY OF WAIKATO, August 13, 2012.
[14] Larose, D., "Discovering Knowledge in Data: An Introduction to Data Mining." New Jersey: John Wiley & Sons, Inc, 2005.
[15] Quinlan Ross J.,"C4.5: Programs for Machine Learning". Morgan Kaufmann Publishers, San Mateo, CA.1993.
[16] Qeethara Kadhim Al-Shayea, "Artificial Neural Networks in Medical Diagnosis", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011.
[17] Two Crows Corporation,"Introduction to Data Mining and Knowledge Discovery. Third Edition." Two Crows Corporation. 10500 Falls Road, Potomac, USA, 2005.
[18] George John and Pat Langley. "Estimating continuous distributions in Bayesian classifiers" In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, pages 338–345, 1995.