



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 3, March 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Heart Disease Detection using Data Mining

Ritesh Maurya, Prof. Manisha Pathak

M.Tech. Scholar, Department of Computer Science & Engineering NITM, Gwalior, MP, India

Assistant Professor, Department of Computer Science & Engineering NITM, Gwalior, MP, India

ABSTRACT: Data Mining is an interdisciplinary study. It can be used in Machine Learning, High Performance Computing, Databases, and Visualization. Data mining in healthcare is a promising new area of research. Data mining and machine learning essentially depends on classification. Data mining in healthcare can be used for various purposes. Many researches are going on various classifiers and feature selection techniques. The classification techniques in the data mining can be applied to the healthcare dataset in order to make valuable predictions and important conclusions. Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. It is one of the predictive modeling approaches used in statistics, data mining and machine learning. In this paper, Simple Cart with Filtered Classifier algorithm and PART are used for the generation of rules for taking the decision for the actual causes of the heart disease. From the result it can be seen that the proposed work is much better than existing techniques.

KEYWORDS: Data mining, KDD, PART (Projective Adaptive Resonance Theory), Machine Learning, WEKA

I. INTRODUCTION

Data Mining is the procedure of examine data from specific summarizing and perspectives the final outcome as helpful information. It has been described as "the non trivial procedure of new, potentially helpful and ultimately conceivable patterns and identifying valid in data" [1].

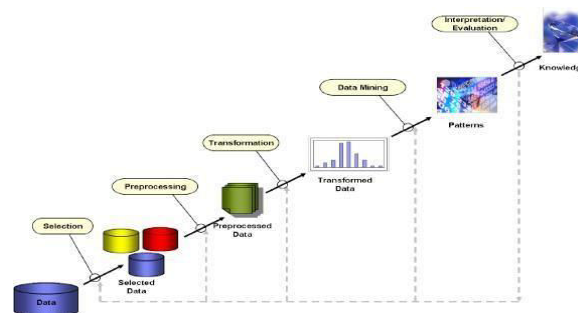


Figure.1 Steps of the KDD Process

The word "Knowledge" in KDD alludes to the revelation of examples which are extricated from the handled information. An example is an articulation depicting certainties in a subset of the information. In this manner, the distinction amongst

KDD and data mining is that "KDD alludes to the general procedure of finding learning from data while data mining alludes to utilization of calculations for extricating designs from data without the extra steps of the KDD procedure." Knowledge" in KDD alludes to the disclosure of examples which are extricated from the handled information. An example is an articulation portraying certainties in a subset of the information. In this manner, the contrast amongst KDD and information mining is that "KDD alludes to the general procedure of finding learning from information while information mining alludes to utilization of calculations for extricating designs from information without the extra strides of the KDD procedure." [2].



A. METHODOLOGIES OF DATA MINING [3]

a. Neural Network- Neural Network or an artificial neural network is a biological system that detects patterns and makes predictions. The greatest breakthroughs in neural network in recent years are in their application to real world problems like customer response prediction, fraud detection etc. Data mining techniques such as neural networks are able to model the relationships that exist in data collections and can therefore be used for increasing business intelligence across a variety of business applications [3].

b. Decision Trees- A decision tree is a flow chart like structure where each node denotes a test on an attribute value, each branch represents an outcome of the test and tree leaves represent classes or class distribution. A decision tree is a predictive model most often used for classification. Decision trees partition the input space into cells where each cell belongs to one class.

c. Genetic Algorithm- Genetic Algorithm is basically used as a problem solving strategy in order to provide with an optimal solution. They are the best way to solve the problem for which little is known. They will work well in any search space because they form a very general algorithm. The only thing to be known is what the particular situation is where the solution performs very well, and a genetic algorithm will generate a high quality solution.

d. CART Algorithm- It stands for Classification and Regression Trees. It was introduced by Breiman in 1984. It builds both classifications and regression trees. The classification tree construction by CART is based on binary splitting of the attributes CART also based on Hunt's algorithm and can be implemented serially.[4]

II. LITERATURE SURVEY

Cheng-Mei Chen and Chien Yen Hsu, et al. (2011) they proposed survival prediction model for liver cancer using data mining techniques. They collect dataset from the medical data center in North Taiwan between the years 2004 and 2008. They extract nine variables to liver cancer survival. Artificial Neural Network(ANN) and Class and Regression Tree(CART) were used prediction model. The model was tested under three conditions: One Variable (Clinical Stage), Six significant variable and all nine variable (Both significant and Non significant). The results shows that ANN model with nine inputs gives 0.915% accuracy, 0.87% sensitivity and 0.88% specificity respectively. Finally they conclude that ANN model gives more accuracy than the CART model. [5]

Ratnamala Kiruba.H et al. (2014) from her paper, she used intelligent agent based system to hike a precise and accurate of diagnosis system. C4.5 decision tree algorithm and Random tree algorithm are used to predict. Two different types of liver disease disorder dataset are combined and predict the accuracy of the disease. And then conclude these both algorithms gives very good accuracy for diagnosing liver disease disorder. [6]

Dhamodharn.S et al. (2014) in his paper he reviewed the classification technique algorithm in data mining techniques for liver disease disorder. Particularly, compared two decision tree algorithms that is FT growth and Naïve Bayes and found out which algorithm gives better accuracy. From his study he said Naïve Bayes is better than FT growth algorithm with the use of machine learning because, Naïve Bayes (75.54%) gives more accuracy than FT growth algorithm (72.66) using WEKA Tool. This comparison happened among 29 datasets with 12 different attributes. [7]

Rajeswari.P, Sophia Reena.G et al. (2010) in his paper, said BUPA liver disorder dataset for early diagnosis the disease. Classification technique algorithms such as Ft tree Naïve Baiyes and Kstar are used to predict the liver disease disorder with evaluate using 10-fold cross validation. Then the results which got from using these algorithms are compared. Generally, accuracy and time taken to build the algorithms are compared and finally she said that from comparing all those algorithms, FT tree gives better role for increasing the accuracy of the dataset in classification technique algorithm. [8]

Aman Kumar sharma, Suruchi, Sahani et al. (2015) in their research they are conducted an experiment in weka environment by using four Decision tree algorithms namely ID3,C4.5,Simple CART and alternating decision tree on



the student dataset and later these four algorithms were compared in terms of the classification accuracy. According to the simulation results, the C4.5 algorithm out performs the ID3, CART and ADTree in terms of Classification accuracy. [9]

Asli calis, Ahmet Boyaci, Kasim Baynal et al. (2015) are discussed in their research about the CART Algorithm has the nature of being the continuation of the decision tree of Morgan and sonquist titled AID and was proposed by Breiman and his team in 1984. The CART algorithm accepts the numeric and the nominal data types as input and predicted variables, it can be used as a solution in classification and regression problems. CART decision tree algorithm As a unique dual from divided into a structure as branching criteria, CART tree is benefits from GiniIndex, without any stopping rule at the phase of its structuring, it is continually divided and grows. In the state where a new branching will not be realized, a cutting out from top in the direction of root is started. The most successful decision tree is subjected to assessment with test data independently selected after each cutting offs and efforts are made to make determinations. [10]

III. PROPOSE WORK

PART:

PART stands for Projective Adaptive Resonance Theory. The input for PART algorithm is the vigilance and distance parameters. Combines the divide-and-conquer strategy with separate-and conquer strategy of rule learning

1. Build a partial decision tree on the current set of instances
2. Create a rule from the decision tree
 - The leaf with the largest coverage is made into a rule
3. Discarded the decision tree
4. Remove the instances covered by the rule
5. Go to step one

Problem Statement:

In the existing technique, they used Simple CART Algorithm which may have unbalanced tree. There is increment or decrement in the complexity of tree when the number of samples changes. The variables and principles are changed then it is the reason of the alterations in the samples of learning by eliminating them. Simple CART is useful for using categorical variables which can handle outliers without difficulty. This method is not suitable for linear structure and create large tree for efficient result. JRip method used for generating final decision rules and it took more time in generating the rules which affects the overall performance of the process.

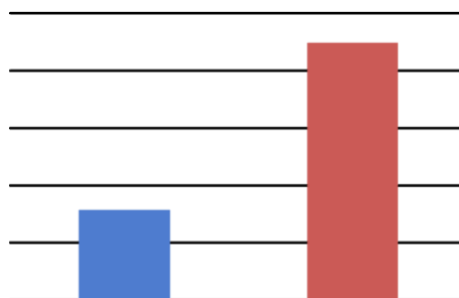


Fig.2 Correctly Classified Instances

Proposed Statement:

In the proposed work, initially load the dataset which should be in csv form. Apply pre-processing over the data and convert the data into nominal form. Heart disease dataset used here to perform the exact reason of the heart disease. Age, chestpain, rest_bpress, blood_sugar, rest_electro, max_heart_rate, exercise_angina and disease are used as an attributes for theclassification of the causes. Apply Simple Cart with Filtered Classifier algorithm for the taking the decisions on the applied dataset. Lastly, PART used for the interpretation of rules and then various paremeters are



calculated for the effectiveness of the proposed technique. PART is the very efficient technique for generating accurate rules for each predictor and then it choose one rule with minimum error.

Proposed Algorithm:

- Step:1 Load the Heart disease dataset in csv format
- Step:2 Perform preprocessing over the dataset
- Step:3 Conversion of data into nominal values
- Step:4 Apply Simple Cart with Filtered Classifier for decision
- Step:5 Convert data from nominal to binary
- Step:6 Generate rules for further decisions using PART
- Step:7 For every predictor, calculate target class frequency
- Step:8 If class is frequent
- Step:9 Evaluate rule's error
- Step:10 Select the predictor of minimum error
- Step:11 Calculate the accuracy of the rules

IV. RESULT ANAYSIS

WEKA used in the simulation of the proposed work. In the result section, the dataset used is Heart Disease dataset which includes 76 attributes, but only 14 attributes are taken into consideration. The Cleveland database has been used in this work for evaluating the existence of patient's heart disease.

Performance of Classifier-The classifier output has been generated after applying dataset by taking 8 attributes and 206 instances. It include the dataset detail examination such as information about total instances, classified and unclassified instances, classification accuracy measures, detailed accuracy measures and confusion matrix etc.

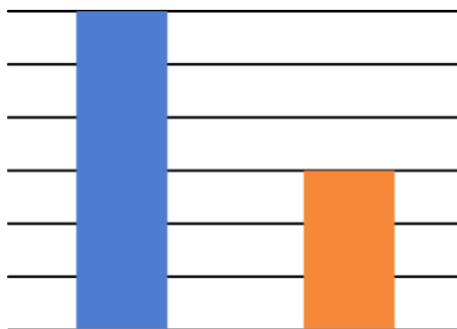


Fig.3 Time taken to build model

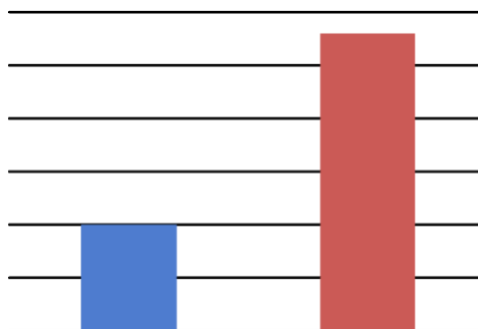


Fig.4 Precision Value

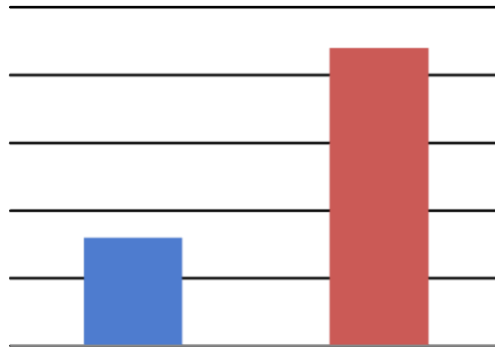


Fig.5 Recall Value

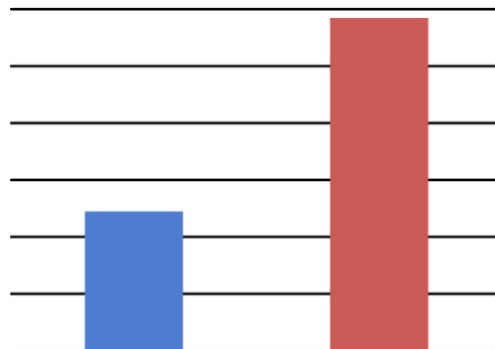


Fig.6 Correctly Classified Instances

Rules Interpretation- Over the dataset, using Simple Cart algorithms by which several rules are produced. These rules are useful for the identification of the actual reason of heart disease. The concluding result is useful by considering the rules produced by PART.

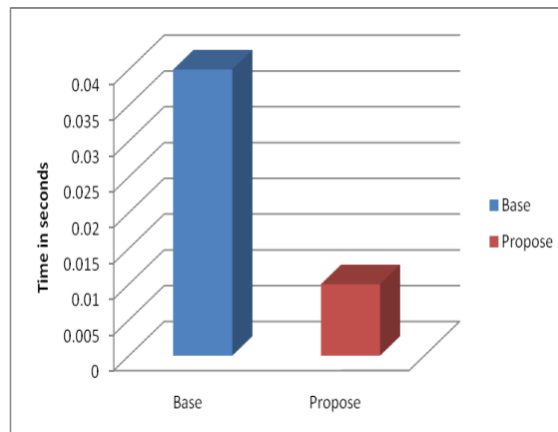


Fig.7 Time taken to build model

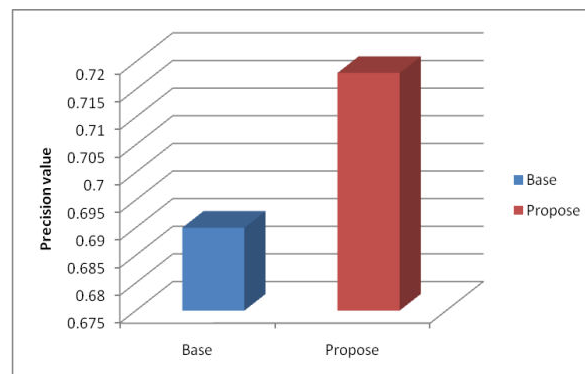


Fig.8 Precision Value

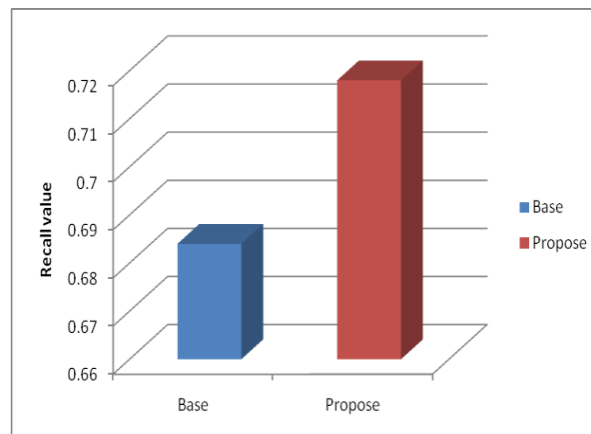


Fig.9 Recall Value

V. CONCLUSION

Data mining is the process of extracting the hidden predictive model from large databases. It has various methods and algorithms. Classification is a supervised method, which builds a model for predicting the new instances. Decision tree classifiers obtain similar or better accuracy when compared with other classification methods. A number of data mining techniques have already been done to improve the performance like Regression, Genetic algorithm, Bays classification, k-means clustering, associate rules, prediction etc. In order to simplify the classification rules and tree, combining Simple Cart with Filtered Classifier algorithm and PART to conduct the attribute and rule reduction on the classification rules of decision tree. The experiment, which compares the Simple CART algorithm with the improved algorithm, shows that the improved algorithm has much better classification efficiency with achieving a simple and efficient classification rule set at the same time.

REFERENCES

1. Mohammadian, M., "Intelligent Agents for Data Mining and Information Retrieval," Hershey, PA Idea Group Publishing, 2004
2. Wang, J., "Data mining: Opportunities and challenges," Idea Group Publishing, September, 2003.
3. Nikita Jain, Vishal Srivastava "DATA MINING TECHNIQUES: A SURVEY PAPER" IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308.
4. Himani Sharma, Sunil Kumar "A Survey on Decision Tree Algorithms of Classification in Data Mining" International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013).
5. Cheng-Mei Chen, Chien-Yeh Hsu, Cheng-Mei Chen and Chien-Yeh Hsu," Prediction of Survival in Patients with Liver Cancer using Artificial Neural Networks and Classification and Regression Trees" Seventh International Conference on Natural Computation 2011.



6. H.Ratnamala Kiruba and Dr G. Tholkappia arasu, “ An Intelligent Agent based Framework for Liver Disorder Diagnosis Using Artificial Intelligence Techniques”, Journal of Theoretical and Applied Information Technology, Vol.69 No.1, 10th November 2014.
7. S.Dhamodharan, “Liver Disease Prediction Using Bayesian Classification”, 4th National Conference on Advanced Computing, Applications & Technologies, Special Issue, May 2014.
P.Rajeswari and G.Sophia Reena, “ Analysis of Liver Disorder Using data mining AAlgorithms”, Global Journal of Computer Science and Technology, Volume.10 issue 14(version 1.00 November 2010 PP 48- 52.
8. Thiptanawat Phonghwattana, Worrawat Engchuan "Clustering-based Multi-class Classification of Complex Disease" IEEE-978-1-4799- 6049-1/15-2015.
9. Kuizhi Mei, Jinye Peng, Ling Gao, Naiquan Zheng Jianping Fan "Hierarchical Classification of Large-scale patient Records for automatic Treatment Stratification "IEEE Journal of Biomedical and Health Informatics., VOL.19.NO.4,JULY 2015.



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details