World Journal of Engineering Research and Technology

WJERT

PREDICTION OF HEART DISEASES USING REPTREE AND ADTREE DATA MINING TECHNIQUES

[1]*Er. Ranbir Singh and [2]Dr. Harmeet Singh

[1]Research Scholar, Department of Computer Science & Engineering, SBBS University, Jalandhar.

[2]Assistant Professor, Department of Computer Science & Engineering, SBBS University, Jalandhar.

*Corresponding Author

Er. Ranbir Singh

Research Scholar,

Department of Computer

Science & Engineering,

SBBS University,

Jalandhar.

ABSTRACT

There are several techniques that are used in data mining, each one having advantages but also disadvantages. To find out which one is most appropriate for our case, when we want to use our databases in a decision-make process we need to have information about our data business and data mining techniques. Alternatively we can try them all and find out which one is the best in our case. The dataset used in this research is based on heart disease obtained different hospitals. This report is based on the findings the reasons to infect from heart disease. As we look at Data Mining tools, we see that there are different algorithms used for creating a decision making (or predictive analysis) system. There are algorithms for creating decision trees such as REPTree and ADTree algorithms for determining performance.

KEYWORDS: REPTree, ADTree, Weka.

INTRODUCTION

The goal of this research is to look at one particular decision tree algorithm called classification algorithm and how it can be used with data mining for heart disease dataset. The purpose is to manipulate vast amounts of data and transform it into information that can be used to make a decision. Decision tree a data mining technique which are REPTree and

ADTree are scalable and fast and are for data streams monitoring from omnipresent devices such as computers, palmtops etc.

Data Mining

An analytical process which is developed to examine data in form of patterns which are consistent is known as data mining. From ages only the physical excerption of patterns from data is going. The data collection, storage and manipulations has increased by the accretion and prevalence computer technology. Due to the grown size as well as complexity of datasets the direct manual analysis has amplified with indirect and automatic processing of data. These are various methods as clustering, neural networks, other genetic algorithms, vector support machines and decision trees which are applied to data with aim of not hiding the patterns which are hidden. Data mining is sometimes turns as knowledge discovery and its tools are here to predict behaviours and future trends making proactive business and knowledge driven decisions. These sophisticated data analysis tools used by data mining are to discover not the known, patterns and relationships of them in large datasets. Various models as statistical and mathematical along other machine learning methods can be involved as tools of data mining. These are basically the algorithms which improves their performance automatically by experience such as neural networks and decision trees.

Techniques of Data Mining

The three main techniques of data mining are

1. Clustering- Clustering is the process in which the objects of similar kind are grouped together into various classes which are termed as clusters. Various customer groups are discovered by the cluster analysis and the characteristics of each group is also analysed. This is the common technique used for market analysis.

2. Classification- The technique involved in predicting certain outcome based on any given input is known as classification. This approach involves certain processes of mining which are made to discover rules which are used to define the sub processes of the technique which bare model building and predicting. In this terms are belonged to class or particular subset of data.

3. Association Analysis- The analysis which shows the association rules discovery giving value conditions of attribute which are constantly occurred in a given dataset. This analysis is very much popular in transaction data analysis and market basket.

Decision trees

There are variety of algorithms being used in classification technique. One if these is the decision tree approach. To represent both the regression models and classifiers decision tree in the state of predicative model is used. Decision tree basically us the hierarchal model of decisions and their consequences. The structure of decision tree includes branch, root node and leaf node. Attributes test is denoted on each interval node, the test outcome is denoted by branch and class labels are shown by leaf node. The topmost node is the root node of the tree. The tree learning is done by dividing the source into set which are generally based on a test of attribute value. The top down approach of decision tree sets an 5 example of greedy algorithm. Apart from this bottom-up approach is also common these days.

RepTree

RepTree uses the regression tree logic and creates multiple trees in different iterations. After that it selects best one from all generated trees. That will be considered as the representative. In pruning the tree the measure used is the mean square error on the predictions made by the tree. Basically Reduced Error Pruning Tree ("REPT") is fast decision tree learning and it builds a decision tree based on the information gain or reducing the variance. REP Tree is a fast decision tree learner which builds a decision/regression tree using information gain as the splitting criterion, and prunes it using reduced error pruning. It only sorts values for numeric attributes once. Missing values are dealt with using C4.5's method of using fractional instances. The example of REP Tree algorithm is applied on UCI repository and the confusion matrix is generated for class gender having six possible values.

ADTree

An alternating decision tree (ADTree) is a machine learning method for classification. It generalizes decision trees and has connections to boosting.

An ADTree consists of an alternation of decision nodes, which specify a predicate condition, and prediction nodes, which contain a single number. An instance is classified by an ADTree by following all paths for which all decision nodes are true, and summing any prediction nodes that are traversed.

RESULTS

As referenced before, our dataset was gotten from different hospitals
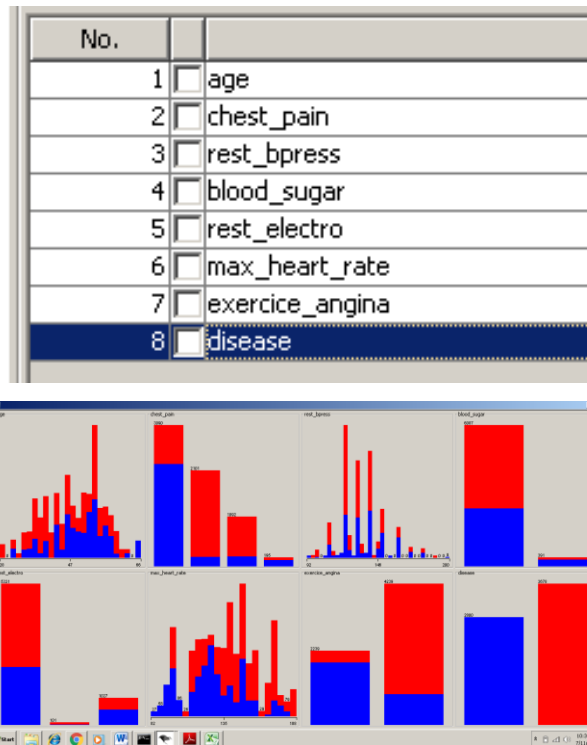
Attributes of Dataset

Fig. 1: Shows that in which category the symptoms of the disease are the highest.

Table 1: Classification Accuracy Values of Algorithms.

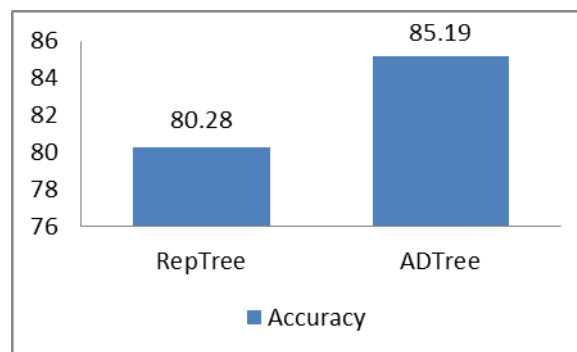|	RepTree	ADTree
Accuracy	80.28	85.19

Fig. 2: Shows the accuracy levels for all classifiers. It shows that ADTree has better accuracy whereas the REPTree shows poor accuracy levels.

Table 2: Classification Error Rate Values of Algorithms.

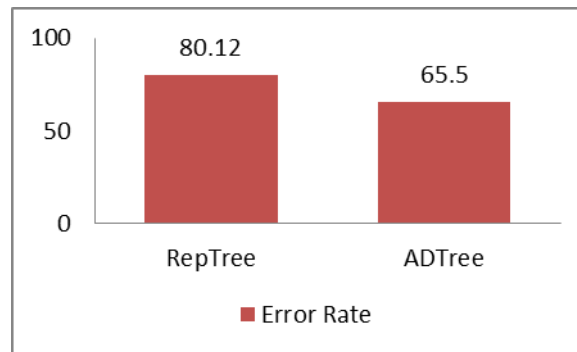|	RepTree	ADTree
Error Rate	80.12	65.5

Fig. 3: Represents Error Rate values for the chosen classifiers. The K values show the better results for ADTree over other classifier.

CONCLUSION

In this exploration work, three classifiers ADtree and REPTree have been utilized for experimentation on WEKA information mining instrument. These classifiers have been contrasted all together with decide the arrangement precision dependent on execution estimates like accurately characterized occurrence, all out an ideal opportunity to assemble model, Error Rate and Accuracy esteems. In this exploration work, the exactness estimates like ADtree has shows the better result in Acuuracy and arror rate as compare to REPTree.

REFERENCES

1. Dhara B. Mehta, Nirali C. Varnagar, "Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining", Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8, 2019.

2. Chitra Jalota, Rashmi Agrawal, "Analysis of Educational Data Mining using Classification", 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th, Feb 2019.

3. Esther Gyimah, Delali Kwasi Dake, "Using Decision Tree Classification Algorithm to Predict Learner Typologies for Project-Based Learning", IEEE, 2019.

4. Shanjida Khan Maliha, comRomana Rahman Ema, comSimanta Kumar Ghosh, "Cancer Disease Prediction Using Naive Bayes,K- Nearest Neighbor and J48 algorithm", 10th ICCCNT 2019 July 6-8, IIT - Kanpur Kanpur, India, 2019.

5. Azwa Abdul Aziz, Andrew Starkey, Elissa Nadia Madi, "Predicting Supervise Machine Learning Performances for sentiment analysis using contextual-based approaches", IEEE, 2019.

6. Muhammad Hammad, Haris Anwar, "Sentiment Analysis of Sindhi Tweets Dataset using Supervised Machine Learning Techniques", IEEE, 2019.