

# Enhance Decision Tree Techniques on Mobile Environment in Data Mining

Parminder Singh, Amarjit Kaur

**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. This research is based on the findings maximum use of mobile service. The results in this report are based on data from mobile service related. 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 ID3 and CART along with algorithms for determining known nearest neighbor or clustering when working on classification. The goal of this research is to look at one particular decision tree algorithm called enhanced algorithm and how it can be used with data mining for mobile service. The purpose is to manipulate vast amounts of data and transform it into information that can be used to make a decision.

**Keywords:** techniques, advantages, appropriate (or predictive analysis), CART, ID3, Alternatively

## I. INTRODUCTION

There are varieties of algorithms being used in classification technique. One if these are the decision tree approach. To represent both the regression models and classifiers decision tree in the state of predicative model is used. Decision trees basically use the hierarchical 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 example of greedy algorithm. Apart from this bottom-up approach is also common these days.

## II. RELATED WORK

- (a) Qiang Yang *et al* (2007): Qiang Yang have discovered a unique algorithm that suggest actions which changes customers undesired status to a desired status such as from attractors to loyal. They have achieved this by maximizing the net profit which is expected and this is their main objective function that is the expected net profit.

**Revised Version Manuscript Received on April 03, 2017.**

**Er. Parminder Singh**, Assistant Professor & Head, Department of Computer Science and Engineering, Ramgarhia Institute of Engineering & Technology, Phagwara (Punjab)-144401, India. E-mail: [singh.parminder9@gmail.com](mailto:singh.parminder9@gmail.com)

**Er. Amarjit Kaur**, M.Tech Scholar, Department of Computer Science and Engineering, Ramgarhia Institute of Engineering & Technology, Phagwara (Punjab)-144401, India. E-mail: [amarjitk664@gmail.com](mailto:amarjitk664@gmail.com)

The main purpose for designing the algorithm was that certain technique which were applied to various problems of industry such as customer relationship management requires human experts for post processing the manually generated knowledge.

- (b) Mingquan Ye *et al* (2013): In this paper an innovative multi-level rough set model (MLRs) which is based on attribute value taxonomies and a program of full sub tree generalization is presented. The researchers have compared the results of MLRs with that of the Pawlak's rough set model.
- (c) Gilad Katz *et al* (2014): Gilad Katz have developed a method named confDtree (confidence-based decision tree) which can be used for the three drawbacks of decision trees. According to the researchers there are these problems which effects decision trees which are performance reduction while dealing with the small training set; criteria of decision tree is very solid and exact; and that a single uncharacteristic attribute sometimes results in derailing of the process of classification

## III. PROBLEM FORMULATION

1. Enhanced decision tree algorithm which will work on large scale high dimensional dataset- there is a problem of data mining in the classification of large datasets. There is no such algorithm stated that performs well in this problem. An algorithm can be made with certain split selection methods involved from the literature which includes algorithms like C4.5 and Jlem48.
2. Enhancement in the efficiency of decision tree construction- various pruning techniques are proposed which can help in the improvement of decision tree construction.
3. Analysis between the computation times- the computation time can be reduced by making alterations in the number of node and leaves. Lesser the no of nodes lesser will be the computation time of the algorithm.
4. Reducing present error rate- the errors rates produced by a predictive model can be reduced by the algorithms. Basically error rate is one minus accuracy.

## IV. FUTURE SCOPE

This proposed technology will help the mobile service providers to analyse the amount of services present at any location and will tell that which services are more used by users. This will include all the services which are generally provided by the mobile networks and with this they will get.



To know that which service has less frequency at which location. By this analysis the service providers will be able to have required services at all various possible locations. The proposed technology will use an enhanced algorithm of decision trees which is an algorithm of classification approach and will perform a predictive analysis to analyse the services at various locations.

## V. OBJECTIVE

The main objectives for the research focuses on the analysis of data in mobile environment with the help of decision tree technique are as follows

1. To propose an idea of complex activities which study the continuously changing behaviour patterns of mobile users. The activity is made as a collection of requests by the services, location movement, the location and service co- occurrence or the interlacing of all.
  2. To analyse different activities which may exhibit dependencies that affect user behaviours.
  3. To propose new methods for analysis of the services present at different locations.
  4. To analyse users activities which will help service providers to provide high quality of services to user at right place and right time.
- To increase accuracy by reducing error rate and reduce response time

## VI. RESULTS AND DISCUSSION

### A. Dataset

To evaluate the performance of our algorithm we have tested it on banking dataset which explains the types of banking services used by various customers[5]. It consists of 15 attributes namely Qualified for rebate, Rate of interest, Interest compound for period, Bonus percentage, With drawl restriction, Interest on tax, Loan/Advance against deposit, payment of return, Nomination facility, Premature closure, Payment rule, Transferability, Minimal deposit, Maximal deposit, Banking service. It has approximately 15 attributes and 5264 instances.

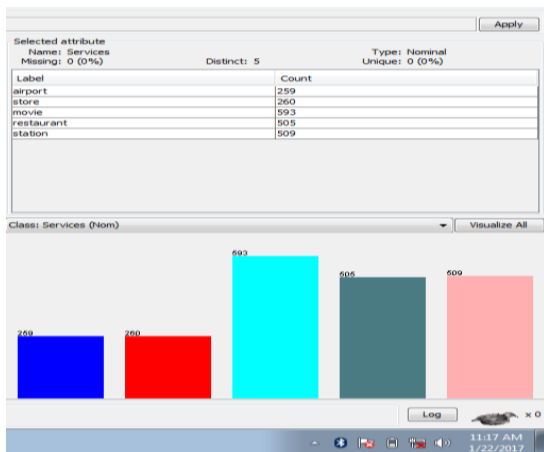


Figure 1. shows the main WEKA Explorer interface with the dataset loaded. The last attribute mobile services is taken as a class attribute by the WEKA . This attribute contains five services .The count of number of instances under each service in the dataset is shown numerically as well as graphically.

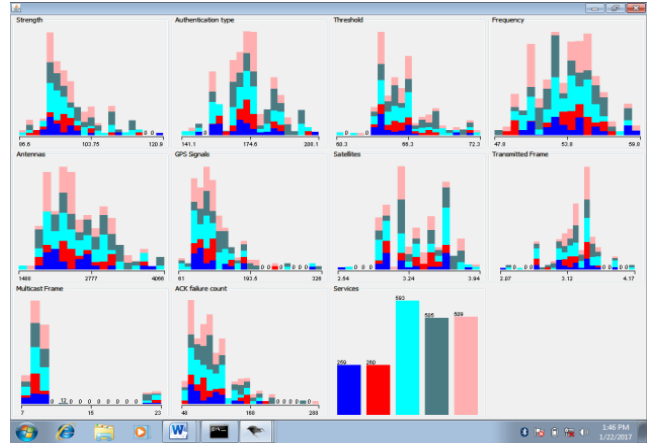


Figure 2. Shows the graphical representation of all the 11 attributes of the mobile dataset. Each bar shows the distribution of four service of class attribute on that particular value.

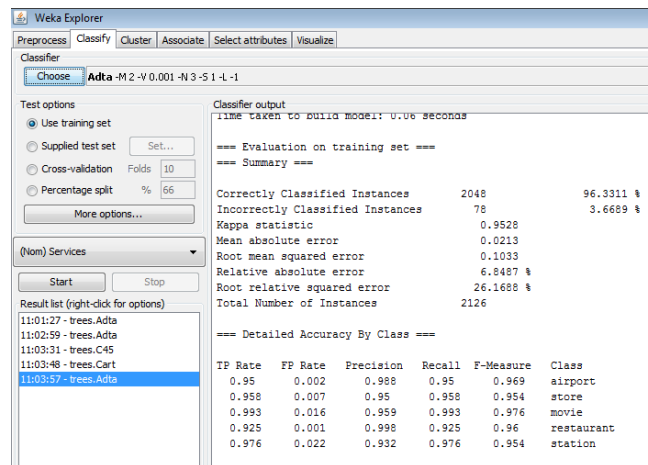


Figure 3. Shows the result of ADTA algorithm when applied on the processed dataset. Results depict the correctly classified instances, incorrect classified instances and error rate. The result window also shows the accuracy of the class.

Table 1 Comparison among ADTA, C4.5 AND CART Algorithms with Correctly Classified Instances, Incorrect Classified Instances and Error Rate on Mobile Dataset

	Correctly Classified Instances	Incorrect Classified Instances	Error Rate
ADTA	96.33	3.66	26.16
C4.5	80.33	19.66	62.32
CART	31.51	68.48	99.01

Table 2. Comparison among ADTA, C4.5 AND CART algorithms with Correctly Classified Instances on Mobile Dataset

	ADTA	C4.5	CART
Correctly Classified Instances	96.33	80.33	31.51

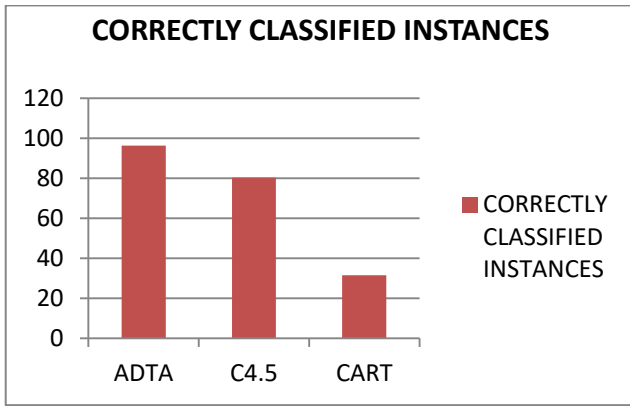


Figure 4. Graphical Representation of Correctly Classified Instances

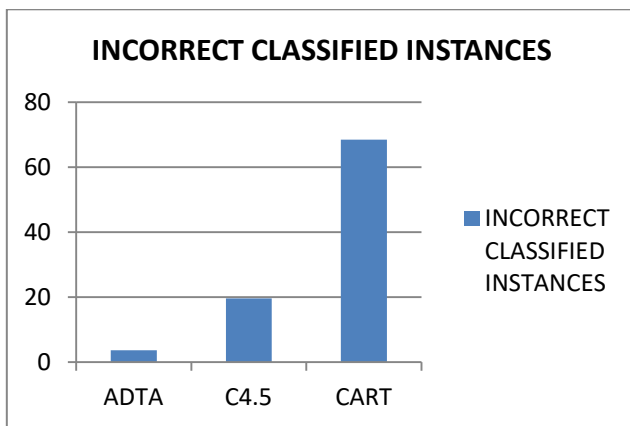


Figure 5. Graphical Representation of Incorrect Classified Instances



Figure 6. Graphical Representation of Incorrect Classified Instances

## VII. CONCLUSION

In this research paper, study is being done on decision tree algorithms. The features of C4.5 and CART decision tree algorithms are combined and a new algorithm ADTA is proposed. The comparison of proposed algorithm is done with the existing algorithms C4.5 and CART on mobile services dataset using WEKA data mining tool. The results by changing the correctly classified instances, incorrect classified instances and error rate values specifies that the proposed method gives better performance than C4.5 and CART by reducing the error rate which signifies that ADTA have high intra similarity and is more accurate. Also the

proposed algorithm can handle large datasets more effectively.

## REFERENCES

1. Margaret H.Dunham,"Data Mining Introductory and Advanced topic", published by person education Delhi, India,[2004].
2. K. Cios, W.Pedrycz, and R. Swiniarski. Data Mining Methods for Knowledge Discovery. Boston: Kluwer Academic Publishers,[1998]
3. Omer Adel Nassar,Dr.Nedhal A.Saiyd,"the integrating between web usage mining and data mining techniques,"5<sup>th</sup> internal conference on computer science and information technology,[2013].
4. Shahida Sulaiman, "Data Mining Technique for Expertise Search in a Special Interest Group Knowledge Portal", 2011 3rd Conference on Data Mining and Optimization (DM O) 28-29 June [2011].
5. Ren Yanna, " The Design of Algorithm for Data Mining System Used for Web Service" ,IEEE [2011] .
6. B.N.Lakshmi,G.H Raghunandhan "A conceptual overview of data mining",IEEE ,Proceeding of the national conference on innovation in emerging technology,pp.27-32,17&18 feb,[2011].
7. G. Sathyadevi "application of CART algorithm hepatitis disease diagnosis", IEEE-International Conference on recent trends in information technology, ICRTIT 2011, June 3-5,[2011].
8. Quinlan J R," Induction of decision tree," Machine Learning, vol.4,no.2,pp.81-106,[1986].
9. Shioh-yang wu, Hsiu-Hao Fan" Activity-based proactive data management in mobile environments IEEE transaction on mobile computing ,vol 9,no.3 March[2010].