

# A Study on Efficiency of Decision Tree and Multi Layer Perceptron to Predict the Customer Churn in Telecommunication using WEKA

S. Babu  
Research Scholar  
Department of CSA,  
SCSVMV University  
Enathur, Kancheepuram

N.R. Ananthanarayanan,  
PhD  
Associate Professor  
Department of CSA,  
SCSVMV University,  
Enathur, Kancheepuram.

V. Ramesh, PhD  
Assistant Professor  
Department of CSA,  
SCSVMV University,  
Enathur, Kancheepuram.

## ABSTRACT

The Data mining is the technique to discover the knowledge which is hidden in the large data sets. It involves with different methods and algorithms to perform efficient analysis over the data sets. The classification is the technique, which is used to mine the data and helps to make the prediction about the future. Different data mining algorithms are available for classification, like C4.5, Simple Cart, NavieBayesen, Logistic Regression and Multi Layer Perceptron based on Artificial Neural Network. The main objective of this paper is to analyse the efficiency of various classification algorithms in terms of performance, accuracy and time complexity. Telecommunication churn dataset is used for the analysis. The obtained results revealed that MLP algorithm outperformed in terms of accuracy and C4.5 algorithm provides better performance in terms of time complexity.

## Keywords

Data mining, Classification algorithm, Decision tree, Telecommunication, Weka.

## 1. INTRODUCTION

Data mining is the process of analyzing the data using variety of tools to discover patterns and the relationship that exists in the data; the same may be used to make valid predictions. Data mining consists of various techniques like Association, Classification, Clustering, and etc. In which, Classification is the technique that, aims to find a model for class attribute, that may be used as a function values for other attributes. The same model can also be used to understand the existing data and to predict how new instances will behave.

There are many classification algorithms has been proposed, but Decision Tree is the one that provide easily readable rules of classification, quick construction of decision trees, and provides better accuracy. Decision tree is a tree, where test on attributes are represented by internal nodes, outcomes of the test are represented by branch nodes and the class labels are represented by leaf node. Very particularly decision tree classifiers provide best accuracy.

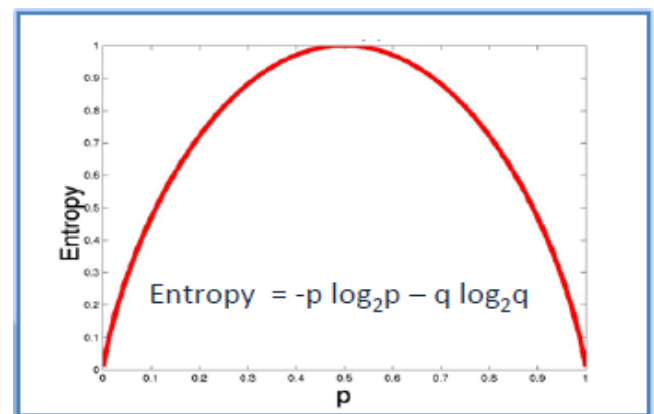
### 1.1 ID3

The algorithm ID3 is a simple algorithm for decision tree developed by Quinlan, 1986. It used top down fashion to build the tree from root node and partitions the data into subsets, which contain the instances with homogenous values.

#### 1.1.1 Entropy

ID3 uses entropy to calculate the homogeneity of a sample [10]. If the instances are homogeneous the entropy is zero and

if the instances are divided equally, then the entropy is one. [10].



$$\text{Entropy} = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$$

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

#### 1.1.2 Information Gain

The information gain is the decrease in entropy that, after the dataset is split based on the attribute. Construction of tree is done by identifying the attribute that has the high information gain.

$$\text{Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X)$$

### 1.2 C4.5

C4.5 is the extension of ID3, developed by the Quinlan, 1993. The Java implementation of C4.5 in weka is the J48 algorithm. C4.5 can handle both discrete and continuous values. Gain ratio is used as splitting criteria. The splitting process will be ceased when the number of instances is less than the certain threshold. Error based pruning will be performed after the tree has been built [9].

### 1.3 CART

CART stands for Classification and Regression Trees developed by Breiman in 1984. It constructs the binary trees, in which each internal node has two edges. The splits are based on towing criteria and tree which is obtained is pruned by cost complexity pruning. The main feature of the CART is, it builds the regression trees whose leaves predict the real number [9].

### 1.4 NB Tree

The Naive Bayesian tree(NBTree – Kohavi, 1996) is a combined one with Naive Bayesian Classification and Decision Tree algorithm. The NBTree works in a very simple manner and in some cases it provides better result. The Naive Bayes is based on the conditional probability of Bayes rule. It used all the attributes in the dataset and analysis them individually even though they are important and independent to each other [11]. The algorithm for learning an NBTree is similar to C4.5.

### 1.5 Logistic Regression

Logistic regression refers to statistical models where the dependent, or outcome variable, is categorical, rather than continuous [8]. The logistic function “maps” or “translates” changes in the values of the continuous or independent variables on the right-hand side of the equation to increasing or decreasing probability of the event modelled by the dependent, or left-hand-side, variable [8]. The main difference of logistic regression with the comparison of other regression model is that the predicted dependent variable is a probability.

### 1.6 Multi Layer Perceptron

Artificial Neural network is the one, which process the information based on biological nervous system. Today large number of real world problems is solved using ANN. The main advantage of ANN is that, it can solve problems that do not have the algorithmic solution [3] and it is specifically configured for the application like data classification. Many types of ANN are there, out of which the simplest one is feedforward neural network. In this method the information moves in forward direction [3]. The ANN method MLP(Multi Layer Perceptron) adopted in this study is an example of feedforward neural network. The MLP maps the input data on to the appropriate output. It has multiple layers with each one connected to next in a directed graph. It has three layers namely input, output and hidden. The hidden layer performs intermediate computation before forwarding the input to the output layer.

## 2. REVIEW OF LITERATURE

### 2.1 [RupaliBhardwaj , Sonia Vatta] [1]

The decision tree algorithms are used for the purpose of the following reasons

1. Decision Tree infer good on the unobserved instance
2. The effective calculations that made by these methods are propositional to the instance that are to be observed
3. The tree generated is easily understood.

The objective of this paper is to generate the decision rules for the CRICKET.Training data set using ID3 algorithm. In this research paper how the data is converted from continuous to discrete has been pointed out. It also pointed out the step by

step procedure, how the ID3 algorithm works to generate the decision tree and decision rule with the above said dataset.

### 2.2 [Gaganjot Kaur, Amit Chhabra] [2]

In this research article, the authors discuss the efficient mining process for the prediction of the diabetics from the patient's medical records. The data set Pima Indians are used to perform the above said process. To predict the j48 algorithms has been used and also new methodology also proposed to improve the accuracy of the classifier. In the author's methodology, weka tool is used as an API for MATLAB for generating the j48 classifiers. Experimental results show that, there was comfortable improvement was found on the proposed methodology. In addition to that the same proposed method output accuracy was also compared with the output accuracy of the other classifiers. By this comparison it is proved that the proposed methodology outperformed.

### 2.3 [O..O. Adeyemo& T. .O Adeyeye, D. Ogunbiyi] [3]

The main objective of this research article is to compare the performance accuracy of the Decision Tree algorithms like ID3, C4.5 against the Multilayer Perceptron (MLP) in the prediction of Typhoid fever. The data set which is collected from the Nigerian hospital was used. To analyze the performance of the above said algorithms, the weka tool is used. The collected data is partitioned into training set and test set. The training set was used to identify the relationship exit between attributes for prediction. The test data was used to analyze the performance of the prediction model. The error performance measures like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Relative Absolute Error (RAE) error performance measures are also considered to evaluate the performance and from that it is concluded that MLP has outperformed.

## 3. METHODOLOGY

The focus of this paper was designed in two phases.

**Phase 1:** During this phase various classifiers was executed on telecommunication dataset, to identify the best classifier for customer churn prediction.

**Phase 2:** During this phase the comparative study was done between MLP(ANN) and the efficient classifier from phase 1.

### 3.1 Data Description

The Telecommunication Data set is used in our work to predict the customer churn. The dataset contains 10 attributes like Phone no, state, Day calls, Day charge, Evening calls, Evening Charges, Night calls, Night Charges, Customer Services calls, Churn status. Every record in the dataset is related to one customer.

### 3.2 Data Preprocessing

Data in the dataset was preprocessed before performing the evaluation of the algorithms.

### 3.3 Data Integration

As the data collected was from different data store, they are merged together in Microsoft Excel 2007.

### 3.4 Data Transformation

In this stage the data selected was transformed to the one which is acceptable by weka. The data file was saved as the CSV(Comma Separated Value) file using the excel. As a next

step, it was converted to ARFF(Attribute Relation File Format) file in the weka.

## 4. IMPLEMENTATION

### 4.1 Phase I

In this paper, the data mining tool weka was used. The dataset was loaded into the weka explorer. The classify tab enables the user to apply various data mining algorithms like, classification and regression on the dataset. To evaluate the performance of the algorithm various measures like, confusion matrix, evaluation time, correctly classified instances, in correctly classified instances and the error performance measures like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Relative Absolute Error (RAE) error performance measures are all calculated by the tool. The algorithms like C4.5, CART, NB TREE, Logistic Regression and MLP are executed on the data set to predict the customer churn and to evaluate the performance efficiency of these algorithms, the above said measures has been considered.

After applying various algorithms on the data set, performance of the respective models was analyzed. Some measures were used to predict the customer churn in the dataset. The following are the measures that were used to analyze the performance of the classifiers: Correctly Classified instances, incorrectly classified instances, time taken to build model, MSE, RMSE, RAE, RRSE.

Table -1 Classifier Accuracy

Classifier	Correctly Classified instances	Incorrectly Classified Instances	Time taken to build model (in sec)	Accuracy (%)
CART	2850	483	40.85	85.51
J48	3213	120	0.32	96.40
Logistic regression	2888	445	1.78	86.65
NB Tree	3176	157	27	95.29

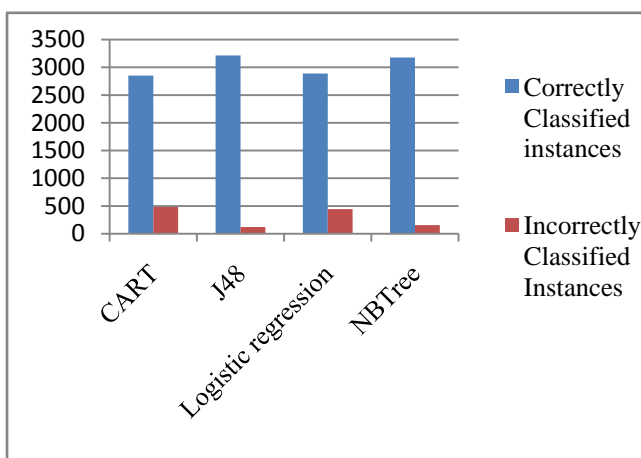


Figure 1: Comparison of Performance Accuracy of various classifiers

The accuracy of the classifier is calculated based on the correctly classified instances. From the Table 1 and Figure 1, it has been clearly identified that the classifier j48 produces better accuracy than the other classifiers. Apart from accuracy, by considering the time taken to build the model, j48 had built the model with the minimal time.

The Table 2 and Figure 2 reveals the output, related to the performance errors derived from each classifiers. The error values like MSE, RMSE, RAE and RRSE are considered. From the result, that j48 is more suitable classifier for prediction of the customer churn in telecommunication with less error values.

Table 2: Performance error table of various classifiers

Classifier	MSE	RMSE	RAE (%)	RRSE (%)
CART	0.2478	0.352	99.94	100.00
J48	0.0529	0.1608	21.34	45.67
Logistic regression	0.1879	0.3068	75.76	87.14
NBTree	0.0846	0.2056	34.10	58.42

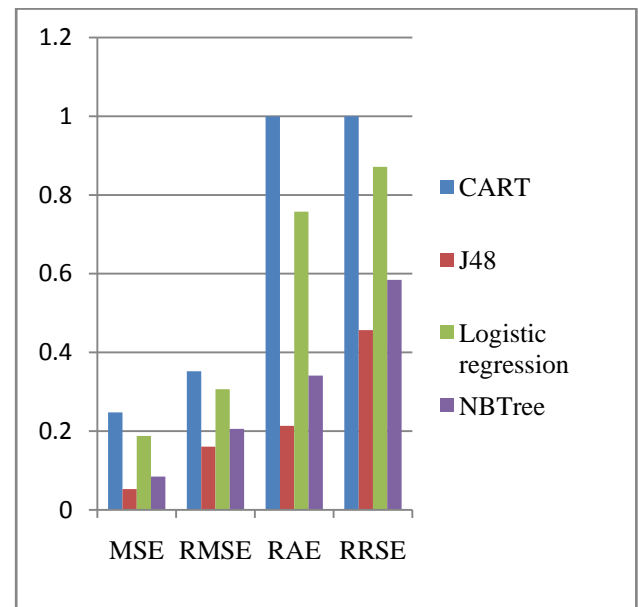


Figure 2: Comparison of performance error value of various classifiers.

### 4.3 Phase II

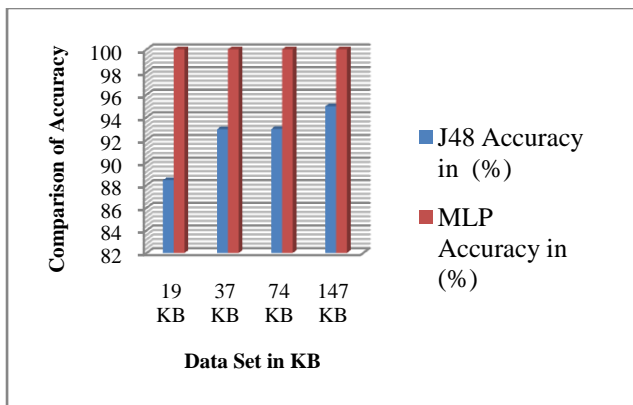
From the Phase I analysis it has been identified that j48 proves to be the best among the other classifier algorithms, in terms of output accuracy as well as performance errors. In Phased II j48 has been compared with MLP(ANN) to identify the best predictor among them. The data on the Table 3 and Figure 3 shows that the MLP can attain the maximum over the output accuracy and proved that MLP is outperformed compared with J48.

**Table 3: Performance Accuracy of J48 and MLP**

Size of the Data Set in (KB)	J48			MLP		
	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy (%)	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy (%)
19	176	23	88.44	199	0	100
37	370	28	92.94	398	0	100
74	740	56	92.96	796	0	100
147	1512	80	94.97	1592	0	100

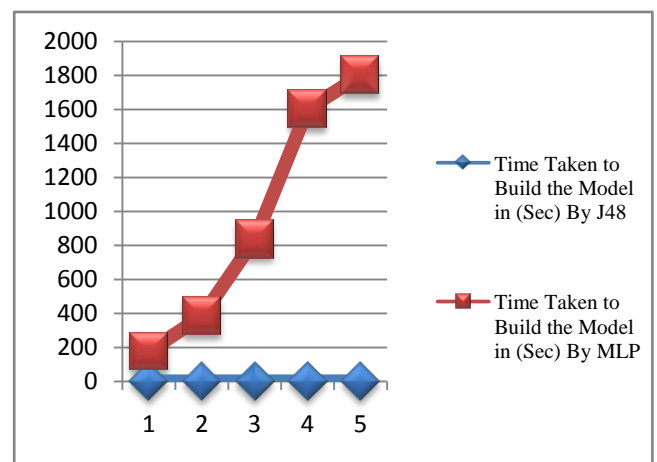
**Table 4: Time and Memory Compatibility of J48 and MLP**

Size of the Data Set in (KB)	J48		MLP	
	Time Taken to Build the Model in (Sec)	Compatibility with Initial JVM Memory Size	Time Taken to Build the Model in (Sec)	Compatibility with Initial JVM Memory Size
19	0.08	Yes	177.54	Yes
37	0.01	Yes	383.21	Yes
74	0.03	Yes	836.66	Yes
147	0.05	Yes	1603.42	Yes
157	0.11	Yes	???	No



**Figure 3: Comparison of Performance Accuracy of J48 and MLP.**

The efficiency of the algorithms not only depends on output accuracy, it also depends on other factors like time taken to execute, amount of memory resources used to execute etc. The data set is fragmented as shown in table 4 and the same was executed on both algorithms to analyze about the time taken and memory utilized to produce the model. As first the data from table 3 and figure 4 depicts that, the time which is taken by the MLP is incomparable with the time taken by J48 to build the model. As second, the MLP can execute at the maximum of 147 KB (No of attributes: 10 and No of Instances:1592) of data set with the initial memory of JVM. Whereas it is also verified that, J48 can execute up to 1912 KB (No of attributes: 10 and No of Instances: 200186) of dataset with the same memory of JVM. In this aspect of analysis, it is proved that the efficiency of J48 is better than MLP. In addition to this from the table3 it is also known that, when the size of the dataset increased the j48 output accuracy also increases.



**Figure 4: Comparison of Time taken to build Model**

## 5. CONCLUSION

In this paper, the various classifier algorithms were studied and the same was executed on the telecommunication data set to predict the customer churn. The paper classified the study into two phases. In the first phase, the algorithms like, J48, CART, NB Tree and Logistic Regression are executed on the dataset to find the efficient classifier. From the above four classifiers, J48 is outperformed both in terms of output accuracy 96.40% and in terms of minimal performance errors. In the second phase the comparison done among j48 and MLP. In consideration of accuracy MLP performed better than J48 with improved accuracy of about 5.03%. In way of time and memory utilization, the efficiency of MLP is incomparable with J48. From this study it is shown that the output accuracy gap of J48 with MLP is very least and able to compromise the gap. But, the time taken and memory utilized by MLP algorithm is too high. In this view it has been concluded that J48 is the efficient for classification and prediction. In addition to this, the study also shows that the techniques of data mining can be used to build effective model for the prediction of customer churn in telecommunication.

## **6. REFERENCES**

- [1] Rupali Bhardwaj, SoniaVatta, “Implementation of ID3 Algorithm”, *International Journal of Advanced Research in Computer Science and Software Engineering*
- [2] Gaganjot Kaur, Amit Chhabra, “Improved J48 Classification Algorithm for the Prediction of Diabetes”, *International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014.*
- [3] O.O. Adeyemo&T.OAdeyeye, D. Ogunbiyi, “Comparative Study of ID3/C4.5 Decision tree and Multilayer Perceptron Algorithms for the Prediction of Typhoid Fever”, *African Journal of Computing & ICT, Vol 8. No. 1 – March, 2015*
- [4] S. Babu, Dr. N. R. Ananthanarayanan, V.Ramesh, “A Survey on Factors Impacting Churn in Telecommunication using Datamining Techniques”, *International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 3, March – 2014.*
- [5] BadrHSSINA, Abdelkarim MERBOUHA, ananeZZIKOURI, Mohammed ERRITALI , “A comparative study of decision tree ID3 and C4.5” *International Journal of Advanced Computer Science and Applications.*
- [6] AnujaPriyama, Abhijeeta, Rahul Guptaa, Anju Ratheeb, and Saurabh Srivastavab , “Comparative Analysis of Decision Tree Classification Algorithms”, *International Journal of Current Engineering and Technology ISSN 2277 - 4106, Available online 1June 2013, Vol.3, No.2 (June 2013)*
- [7] MohsimNadaf, Vidya Kadam, “Data Mining in Telecommunication”, *International Journal on Advanced Computer Theory and Engineering (IJACTE), ISSN (Print) : 2319 – 2526, Volume-2, Issue-3, 2013*
- [8] Andrew H. Karp, “Using Logistic Regression To Predict CustomerRetention”,<http://www.lexjansen.com/nesug/nesug98/solu/p095.pdf>.
- [9] Ms. A. Sivasankari, Mrs. S. Sudarvizhi, S. Radhika Amirtha Bai, “Comparative Study Of Different Clustering And Decision Tree For Data Mining Algorithm “, *International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 2, Issue 3, pp: (221-232), Month: July-September 2014*
- [10] [http://www2.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/c4.5/c4.5\\_prob1.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/c4.5/c4.5_prob1.html)
- [11] <http://www.d.umn.edu/~padhy005/Chapter5.html>