Volume 3, Issue 10, October 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Applying Data Mining Techniques in Telecom Churn Prediction

N.Kamalraj
Department of Computer Technology
Dr. SNS Rajalakshmi College of Arts and Science
Coimbatore, India

Dr.A.Malathi
PG and Research Department of Computer Science
Govt. Arts College
Coimbatore, India

Abstract— In the knowledge management process, the techniques that are applied to discover significant knowledge from a large amount of data is called data mining techniques. In recent years, according to the competition in the telecommunication industry keeping customers has become more important for the industry and finding reasons of churning is challenging. The churn prediction facilitates to examine the mobile operators to identify the risky customer. This paper aims to determine the possible churners using the predictive data mining model. The primary objective of the study is to get the complete investigation about the data analysis in the critical process to precede the successful data mining application. It is used to investigate the data analysis, robust predictive model can be built by discovering the significant churn factors. It also examines in keeping the predictive models to make the mobile operators in order to perform them accurately. J48 decision tree technique and C5.0 classification technique is discussed here. The above techniques used here for the large data sets of the telecommunications industry. The focus of this research formulates the better understanding of churn prediction using data mining techniques. Telecommunication industry can use this approach to customer retention activities within the context of their Customer Relationship Management efforts.

Keywords— Churn Prediction, C4.5, J48, C5.0, See5

I. INTRODUCTION

Many companies are finding the reasons of losing customers, measuring customer loyalty and regaining customers have become very important concepts. So, the companies organize various studies and campaigns to avoid losing their customers rather than to obtain new ones [1]. The telecommunication sector acquires huge amount of data due to rapidly renewable technologies, the increase in the number of subscribers and with value added services [2]. Uncontrolled and very fast expansion of this field cause increasing losses depending on fraud and technical difficulties. Therefore, the developments of new analysis methods have become a must. The data mining technology was first adopted by the telecommunications industry [3]. It is because the telecommunication companies are routinely give and depot tremendous quantity of superiority data, hugely rich customer base and changing rapidly with highly competitive environment [4]. The churn is the process of customer who leaving a company and constructing a model of customer attrition is called as churn analysis. The telecommunication industry routinely provided huge cash incentives for customers to switch carriers, it is because nowadays customer churn is a big problem in the industry [5]. Telecommunications churn – leaving a current company and proceeding to another telecommunications company – nowadays it is still a big deal within companies. The company should know why a customer decides to go for other company is essential in the product-range or services by detecting faults [1] [6]. With help of more stored customer data, why he/she churned is becoming more and more interesting by trying to determine relations. Data mining methods are being often used to discover interesting patterns within data. Classification is one of the most common data mining tasks, with the aim of classifying unknown cases based on a set of known examples into one of possible classes. Considering the telecommunications churn problem, the aim of classification is to learn to predict whether a consumer will move to a different company based on the consumers' data stored within company's database [7]. The churn prediction is a supervised classification problem defined as follows from a machine learning perspective: given a set of data, describing each customer's behavior in the network (attributes) together with the information whether the customer has switched to another company, using the values of this same set of attributes the aim is to predict the future churners [8]. For the purpose of this study, a publicly available telecommunications churn problem dataset was used.

II. RELATED LITRATURE

A. *Data Mining Techniques*

All The procedure of reducing, analysis of patterns, relationships and useful, hidden predictive information from large databases is called as Data. There are four tasks requires in this process association rule learning, regression, clustering and classification. Data mining are of two types: the new rules and patterns that are discovered by the system is called as "discovery oriented" and the user's hypothesis is checked by the system is called as "verification-oriented" [9]. It permitting businesses to be proactive, knowledge-driven decisions and also for predicting the future trends and behavior of the businesses [10].

The techniques adopted by data mining adopted its techniques from many research areas as a multi-disciplinary fields, including database systems, rough sets, statistics, visualization, machine learning, and neural networks [11].

A.A.1 Approach for Statistical

Analysis of cluster, correlation, regression and Bayesian network are the statistical tools that are used in data mining. Using the set of training data, the statistical models were built. Using this model the Regularities, rules, and patterns are then built [11].

A.A.2 Approach for Machine Learning

Conceptual clustering, decision tree induction and inductive concept learning are the machine learning methods that are used in the data mining. From the root to a leaf node, an object's class is decided by a decision tree classification; using the attribute values of the object the branches are chosen. From the set of training data the decision trees are induced and using that tree the classification rules are extracted [11].

A.A.3 Other approaches

Rough sets, visualization and neural networks are the other techniques that are used for the data mining. A set of interlinked nodes called neurons are known as a neural network. The membership of the set is fuzzy, then the set is known as rough set [11].

B. Churn Prediction

Churn customer is one who leaves the existing company and become a customer of another competitor company. The management that was assumed to determine the customer turnover is called as Churn management. (Hadden, Tiwari, Roy and Ruta, 2007) [12]. Churn is defined as a discontinuation of a contract. Reducing churn is important because acquiring new customers is more expensive than retaining existing customers [8]. Churn rate measures the number of things moving out and in during a specific period of time. Although, this term is used in many contexts, it is vastly used in business showing the number of customers leaves the company or stay in. The churn rate is broadly used in telecommunication industry. Customer movement from one provider to another in telecommunication industry is called customer churn and the operator's process to retain profitable customers counted as churn management (Berson, Smith & Thearling, 2000) [13].

The reason for churning of customers and when it will happen is analysis by the data mining and it is predicted whether the customer will churn or not. Extending the packages or offering new incentives for customers to remain who are likely to churn from the existing company by predicting and it will reduce the rate of churn of the company [13]. The providers can improve their services for the churn customers by knowing the reason for churn [9]. There are a few decision tree classification algorithms; C4.5,J48, CART, ID3 and etc.

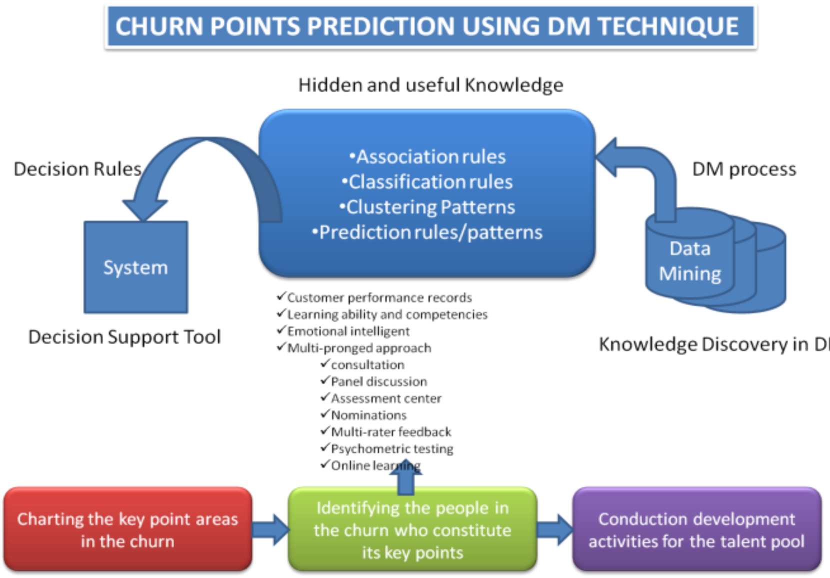

Fig 1: Churn points prediction using DM.

III. DATA SET

In this paper, the data set is from Orange Database and it contains 10 attributes and information about 3333 instances of the company with the details of the customer who churned (left from the company). The attributes are as follows- [14]

TABLE 1: CHURN PREDICTION ATTRIBUTES

S.No.	Attributes
1	Account Number
2	Area Code
3	Voice Mail Service

4	Day Minute(Number of Minutes)
5	Day Calls(Number of calls per day)
6	Day Charge
7	International Minute
8	International Call
9	International Call Charge
10	Churn

IV. RESEARCH METHODOLOGY

There are many classification techniques available for mining the data. The following are techniques are used for the data mining[10].

TABLE 2: DM CLASSIFICATION TECHNOLOGIES

S.No.	Techniques
1	Rule based Classifiers
2	Naïve Based Classifiers
3	Decision Trees
4	Nearest Neighbor
5	Artificial Neural Network
6	Rough Sets
7	Support Vector Machine
8	Fuzzy Logic
9	Genetic Algorithm

In this research, we are using J48 decision tree technique and C5.0 classification technique for prediction of churn customers.

A. J48 Decision Tree Technique

From the available data, using the different attribute values gives the dependent variable (target value) of a new sample by the predictive machine-learning called a decision tree. The attributes are denoted by the internal nodes of a decision tree; in the observed samples, the possible values of these attributes is shown by the branches between the nodes, the classification value (final) of the dependent variable is given by the terminal nodes. Here we are using this type of decision tree for large dataset of telecommunication industry. In the data set, the dependent variable is the attribute that have to be predicted, the values of all other attributes decides the dependent variable value and it is depends on it. The independent variable is the attribute, which predicts the values of the dependent variables.

The simple algorithm is followed by this J48 Decision tree classifier. In the available data set using the attribute value, the decision tree is constructed for assort a new item. It describes the attribute that separates the various instances most clearly, whenever it finds a set of items (training set). The highest information gain is given by classifying the instances and the information about the data instances are represent by this feature. We can allot or predict the target value of the new instance by assuring all the respective attributes and their values [15].

B. C5.0 Classification Technique

An extension of C4.5 algorithm is C5.0 algorithm. C5.0 is the classification algorithm which applies in big data set. C5.0 is better than C4.5 on the efficiency and the memory. C5.0 model works by splitting the sample based on the field that provides the maximum information gain. The C5.0 model can split samples on basis of the biggest information gain field..The sample subset that is get from the former split will be split afterward. The process will continue until the sample subset cannot be split and is usually according to another field. Finally, examine the lowest level split, those sample subsets that don't have remarkable contribution to the model will be rejected [16].

B.A.1 Information Gain

Gain is computed to estimate the gain produced by a split over an attribute [17]

Let S be the sample:

Ci is Class I; i = 1,2,…,m

$I(s1,s2,..,sm)= - \Sigma\, pi \log2\, (pi)$

Si is the no. of samples in class i

$Pi = Si /S$, log2 is the binary logarithm

Let Attribute A have v distinct values.

Entropy = E(A) is

$$\Sigma_{j=1} \{(S1j+S2j+..+Smj)/S\}*I(s1j,..smj)$$

Where Sij is samples in Class i and subset j of Attribute A.

$I(S1j,S2j,..Smj)= - \Sigma\, pij \log2\, (pij)$

$Gain(A)=I(s1,s2,..,sm) - E(A)$

Gain ratio then chooses, from among the tests with at least average gain,

The Gain Ratio= P(A)

$$\sum_{i}^{t} \frac{Si}{S} \log(\frac{Si}{S})$$

Gain Ratio(A) = Gain(A) / P(A)

V. IMPLEMENTATION MODEL

A. J48 Decision Tree Implementation

1. J48 construction is like a flow-chart. A test on an attribute is denoted by internal node. The effect of the test is presented by branch. The class labels or class distribution is represented by leaf nodes.
2. It comprises of two levels: i) The root division examples recursively based on selected attribute for all training examples at the tree construction. ii) The noise or outliers branches are identified and removed by Tree pruning.
3. Decision tree Usage: An unknown sample is classified and the attribute values of the sample are tested by the decision tree.

A.A.1 Algorithm for Decision tree Induction

1. Recursive divide and conquer is used to build the Basic algorithm (a greedy algorithm) Tree in a top-down manner. The root attributes for all the training examples are categorical at first start, depend on selected attributes the examples are partitioned recursively. The basis of a heuristic or statistical measure is used for the selection of Test attributes are selected.
2. To stop the partitioning the conditions should be provided.

A.A.2 Extracting Classification Rules from Trees

1. If-then statement is used to represent the knowledge
2. For each path from root to a leaf one rule is created.
3. A conjunction is formed along a path by each attribute-value pair.
4. The class prediction is containing in the leaf node.
5. It is easier to understand the rules for humans.

B. C5.0 Classification Technique Implementation:

Here C5.0 algorithm is implemented using See5 tool. Like same as the Weka tool, it also produces a confusion matrix of predicted active and predicted churn. In this implementation, the given data set is split into training and testing data set. 80% of the data set ie 2666 instances consists of 2417 active(non-churn) customer records and 249 records of churn customers are taken in training set. Whereas 20% of the data set ie 667 instances consists of 418 records of active customers and 249 non-churn customers records are taken in testing set.

VI. RESULT

A. J48 Implementation Result

A.A.1 Investigation of Performance and Results

Using Weka tool, the classification algorithm is used to perform the experiments on the telecommunication churn data set. From the given data set by using J48 algorithm, the total number of instances is calculated at the first stage. Then the classification accuracy and cost analysis is measured by the experiment at next moment.

A.A.2 Confusion Matrix

In the classification system, actual and predicted classifications information is given by a confusion matrix. With the data of the matrix, the performance of classification systems is judged. For the two class classifier the confusion matrix is shown by the following table. In the context of our study the following gives the meaning for the entries of the confusion matrix:

1. a – when an instance is negative, then it represents the number of correct predictions
2. b –when an instance is positive, then it represents the number of incorrect predictions
3. c –when an instance is negative, then it represents the number of incorrect of predictions
4. d – when an instance is positive, then it represents the number of correct predictions

A.A.2.1 Using Confusion matrix the following standard terms that are derived

1. TP -True positive: When p(positive) is the outcome of the prediction and also actual value, then it is known as true positive (TP).
2. FP-False positive: Whenever the actual value is n(negative), then it is called as false positive(FP)
3. Precision and recall: The fractions of retrieved instances which are relevant is called as Precision, the retrieved relevant instances is denoted by the fraction is called as Recall. Precision and Recall is nothing but it is the measure of relevance.
4. The measure of exactness or quality is called as Precision, a measure of completeness or quantity is called as recall.

An algorithm that returns most of the relevant results is known as High recall. An algorithm that returns the more relevant results than irrelevant is known as High precision.

Precision = $\frac{tp}{tp+fp}$

Recall = $\frac{tp}{tp+fn}$

F-measure=2* Precision* Recall/ Precision+ Recall

Using churn dataset the classification is based on any one of the test options like Cross-Validation, Percentage Split, using training set. The following is the result which is based on the test option training set.

```
Relation:     churn.arff
Instances:    3333
Attributes:   10
              acct
              area
              vmail
              day.mins
              day.calls
              day.charge
              intl.mins
              intl.calls
              intl.charge
              churn
Test mode: evaluate on training data
=== Classifier model (full training set) ===

J48 pruned tree
------------------

day.mins <= 264.4
|   intl.mins <= 13.1: No (2690.0/268.0)
|   intl.mins > 13.1
|   |   intl.mins <= 13.9
|   |   |   intl.mins <= 13.4
|   |   |   |   area <= 415: No (67.0/9.0)
|   |   |   |   area > 415
|   |   |   |   |   acct <= 1847: No (9.0/1.0)
|   |   |   |   |   acct > 1847
|   |   |   |   |   |   acct <= 2435: Yes (6.0/1.0)
|   |   |   |   |   |   acct > 2435: No (5.0/1.0)
|   |   |   intl.mins > 13.4
|   |   |   |   day.mins <= 128.7
|   |   |   |   |   intl.calls <= 2
|   |   |   |   |   |   day.mins <= 109.4: No (2.0)
|   |   |   |   |   |   day.mins > 109.4: Yes (2.0)
|   |   |   |   |   intl.calls > 2: Yes (7.0)
|   |   |   |   day.mins > 128.7
|   |   |   |   |   vmail = yes
|   |   |   |   |   |   intl.calls <= 8: No (17.0/3.0)
|   |   |   |   |   |   intl.calls > 8: Yes (3.0)
|   |   |   |   |   vmail = no
|   |   |   |   |   |   day.mins <= 189.7
|   |   |   |   |   |   |   day.mins <= 158: No (14.0)
|   |   |   |   |   |   |   day.mins > 158
|   |   |   |   |   |   |   |   intl.mins <= 13.6: Yes (4.0)
|   |   |   |   |   |   |   |   intl.mins > 13.6
|   |   |   |   |   |   |   |   |   intl.mins <= 13.7: No (2.0)
|   |   |   |   |   |   |   |   |   intl.mins > 13.7
|   |   |   |   |   |   |   |   |   |   intl.mins <= 13.8
|   |   |   |   |   |   |   |   |   |   |   day.mins <= 175.2: No (2.0)
|   |   |   |   |   |   |   |   |   |   |   day.mins > 175.2: Yes (3.0)
|   |   |   |   |   |   |   |   |   |   intl.mins > 13.8: Yes (2.0)
|   |   |   |   |   |   day.mins > 189.7: No (28.0/7.0)
|   |   intl.mins > 13.9: No (259.0/41.0)
```

day.mins > 264.4
| vmail = yes: No (53.0/6.0)
| vmail = no: Yes (158.0/37.0)

Number of Leaves : 20

Size of the tree : 39

Time taken to build model: 0.46 seconds
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances 2959 88.7789 %
Incorrectly Classified Instances 374 11.2211 %
Kappa statistic 0.3913

Mean absolute error 0.1965
Root mean squared error 0.3135
Relative absolute error 79.2481 %
Root relative squared error 89.0485 %
Total Number of Instances 3333
a b <-- classified as
2812 38 | a = No(non-churners)
336 147 | b = Yes(churners)

The confusion matrix is given by

TABLE 3: CONFUSION MATRIX CLASSIFICATION

a	b	Classified as
2812	38	a= no (Non-Churners)
336	147	b=yes(churners)

Telecommunication Churn Data set have total no. of 3333 instances. Churn class has been chosen randomly from telecommunication dataset. Using J48 algorithm the two values No(Non-churners) and Yes(Churners) can be generated for class churn by the confusion matrix of the given data set.

The confusion matrix that computes the actual and predicted classification is shown in table 3. i.e. for class a 2812 is the total no. of true positives and 336 is total no. of False positive. For class b 38 is the total no. of true positives and 147 is the total no. of false positive. Depends on the test options the confusion matrix will be differ.

TABLE 4:CONFUSION MATRIX CLASSIFICATION-DETAILED ACCURACY BY CLASS

=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.987	0.696	0.893	0.987	0.938	0.673	No
	0.304	0.013	0.795	0.304	0.44	0.673	Yes
Wgt.	0.888	0.597	0.879	0.888	0.866	0.673	
Avg							

B. C5.0 Classification Implementation Result

Using See5/C5.0 tool, the following result are obtained. The confusion matrix and prediction success rate of training dataset and testing dataset are shown in Table 5 and Table 6 respectively.

TABLE 5: CONFUSION MATRIX AND PREDICTION SUCCESS RATE FOR TRAINING DATA

True Class	**Total Number of Samples**	**Predicted Active**	**Predicted Churn**	**Success Percent**
Active	2417	2299	118	95.11
Churn	249	148	101	59.43

TABLE 6: CONFUSION MATRIX AND PREDICTION SUCCESS RATE FOR TESTING DATA

True Class	Total Number of Samples	Predicted Active	Predicted Churn	Success Percent
Active	418	402	16	96.17
Churn	249	167	82	67.06

VII. DISCUSSION OF RESULTS

In this analysis, we experimented with 2 types of classification techniques namely J48 and C5.0 on 3333 instances of telecommunication industry customers in which 2835 were non-churn customers (active customers) and 498 were churn customers. To predict churn customers, the significant customer characteristic is traced out by these two techniques. By comparing the confusion matrix of J48 result and C5.0 algorithm result, J48 yields 11.22 percent classification result on the training data, whereas C5.0 algorithm yields 59.43 percent classification rate on the training data, and 67.06 percent classification rate on the testing data. For the large data sets, the prediction success rate of Churn class is quite high in C5.0 algorithm result than J48 algorithm result. C5.0 algorithm also predicts the significant characteristics of the churn customers than J48 algorithm. And C5.0 algorithm uses less memory than other classification techniques. For reaping higher benefits, a model with higher prediction success rate of Churn class (i.e., C5.0) has to be chosen.

VIII. CONCLUSION

In recent days we find numerous techniques in order to predict the churn. Therefore we need a straightforward method to classify the churners from non-churners. This paper helps to the Customer Relationship Management (CRM) department to know their customers and their behavior for churning. Here a simple model based on DM techniques is introduced. A data set of 3333 instances of 2835 active customers and 498 churn customers with 10 attributes is used to train and test the model. The findings of the result of this paper focuses on the implementation of J48 decision tree, we can predict the churn customers based on actual class using confusion matrix. Using this technique, the number of correctly classified instances is 2959/3333 and the number of incorrectly classified instances is 374/3333. J48 yields 11.22 percent classification result on the training data. Further, the classification technique like C5.0 See5 is also implemented for predicting the churn customers for telecommunication industry. C5.0 algorithm yields 59.43 percent classification rate on the training data, and 67.06 percent classification rate on the testing data. So, C5.0 algorithm is better one to predict the churn customers and their significant characteristics than J48 algorithm and also C5.0 uses less memory than J48. This study helps to predict the future churn of telecom customers to check easily, by formulating intervention strategies based on churn prediction to reduce the lost revenue by increasing customer retention. It is expected that, with a better understanding of these characteristics, telecommunication industry can develop a customized approach to customer retention activities within the context of their Customer Relationship Management efforts.

REFERENCES

[1] Umman Tuğba Şimşek Gürsoy, "Customer churn analysis in telecommunication sector", Istanbul University Journal of the School of Business Administration, Cilt/Vol:39, Sayı/No:1, 2010, 35-49 ISSN: 1303-1732 – www.ifdergisi.org © 2010

[2] Marcin Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts", Expert Systems with Applications, 37 (2010) 4710–4712

[3] Gary M. Weiss, "Data Mining in the Telecommunications Industry", Data Mining and Knowledge Discovery Handbook, 2005 - Springer

[4] Anuj Sharma and Dr. Prabin Kumar Panigrahi, "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services", International Journal of Computer Applications (0975 – 8887) Volume 27– No.11, August 2011

[5] Dr. U. Devi Prasad and S. Madhavi, "Prediction of Churn Behavior of Bank Customers Using Data Mining Tools", Business Intelligence Journal, January - 2012 Vol.5 No.1, ISSN: 1918-2325

[6] V. Yeshwanth , V. Vimal Raj and M. Saravanan, "Evolutionary Churn Prediction in Mobile Networks using Hybrid Learning", Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference, 2011.

[7] Sunita Beniwal, Jitender Arora, "Classification and Feature Selection Techniques in Data Mining", International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 6, August – 2012 ISSN: 2278-0181.

[8] Prof S. Chandrasekhar, "Predicting The Churn In Telecom Industry"'.

[9] Khalida binti Oseman, Sunarti binti Mohd Shukor,Norazrina Abu Haris, Faizin bin Abu Bakar, "Data Mining in Churn Analysis Model for Telecommunication Industry", *Journal of Statistical Modeling and Analytics Vol. 1 No. 19-27, 2010*.

[10] Dr. Rajni Jain, "Introduction to Data Mining Techniques".

[11] Yongjian Fu, "Data Mining: Tasks, Techniques and Applications", Potentials, IEEE, Volume: 16, Issue: 4, Page(s): 18 - 20, 1997.

[12] Shan Jin, Yun Meng, Chunfen Fan, Feng Peng, Qingzhang Chen, "The Research on Applying Data Mining to Telecom Churn Management", 2nd International Conference on Materials, Mechatronics and Automation Lecture Notes in Information Technology, Vol.15, 2012.

[13] Essam Shaaban, Yehia Helmy, Ayman Khedr, Mona Nasr, "A Proposed Churn Prediction Model", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 2, Issue 4, June-July 2012.
[14] Cellular Telecommunications Industry : 1990 to 2006.
[15] Anshul Goyal and Rajni Mehta, "Performance Comparison of Naïve Bayes and J48 Classification Algorithms", International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012).
[16] Tomasz Bujlow, Tahir Riaz, Jens Myrup Pedersen, "A method for classification of network traffic based on C5.0 Machine Learning Algorithm in Section for Networking and Security", International Conference on Computing, Networking and Communications (ICNC), 2012(ieee)
[17] Prof. Nilima Patil, Prof. Rekha Lathi, Prof. Vidya Chitre, "Comparison of C5.0 & CART Classification algorithms using pruning technique", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 1 Issue 4, June – 2012.

