Customer Churn Prediction in Telecom Industry using FURIA and C4.5 algorithm

[1]Dayana Thomas, [2]Dr. Anitha Patil

[1]M. E. Scholar, [2]Professor & Assistant Head of Department
Department of Computer Engineering,
Pillai HOC College of Engineering and Technology, Khalapur Rasayani, Maharashtra 410207

Abstract*: Nowadays, Customer churn prediction in Telecom industry is one of the most important research topics. Customer churn is a condition of switching from one service provider to another by a customer. Telecommunication companies face considerable loss of revenue because some of the customers who are at risk of leaving a company. Customer churn prediction is a foremost feature of contemporary telecom CRM systems. Churn prediction system helps the customer relationship management to retain the customers who are probable to quit. In this paper, we propose a highly sophisticated model using Fuzzy Unordered Rule Induction Algorithm FURIA and C4.5 algorithm to predict customer churn in a telecom industry. FURIA extends the well known RIPPER algorithm, a state-of-the-art rule learner. FURIA learns fuzzy rules instead of conventional rules and unordered rulesets instead of rule lists. Therefore, the proposed model is capable of predicting customers churn behavior well in advance with more accuracy. The FURIA has less execution time compared to the previous research. And also the FURIA outperforms decision tree and other models in terms of accuracy, precision, recall and has the ability to identify the maximum churners for retention campaigns.

***Keywords*: Churn Prediction; CRM (Customer Relationship Management); FURIA (Fuzzy Unordered Rule Induction Algorithm); C4.5 algorithm.**

I. Introduction

Today, telecommunication market all over the world is facing a severe loss of revenue due to fierce competition and loss of potential customers. It is very much challenging and tedious issue to keep the customers intact for a long duration. To survive in the market, telecom operators usually offer a variety of retention policies to attract new customers. This is the major cause of the subscribers leaving one network and moving to another one which suits their needs. The major concern in customer relationship management in telecom companies is the ease with which customers can move to a competitor, a process called “churning”. Churning is a costly process for the company, as it is much cheaper to retain a customer than to acquire a new one. In this situation, the only remedy to overcome such business hazards and to retain in the market, operators are forced to look for alternative ways of using data mining techniques and statistical tools to identify the cause in advance and to take immediate efforts in response. This is possible if the past history of the customers is analyzed systematically.

The telecom industries generate and maintain a large volume of data. They include Billing information, Call detail Data and Network Data. The most common areas of research in telecom databases are broadly classified into 3 types: i) Telecom Fraud Detection ii) Telecom Churn Prediction iii) Network Fault Identification and Isolation. Only the relevant data items which really contribute to the specific analysis must be considered for any study. The researchers would like to focus only on postpaid phones with respect to churn prediction, which is the purpose of this research work.

Churn Prediction is a phenomenon which is used to identify the possible churners in advance before they leave the network. For that, a churn prediction model is made which helps the CRM department to prevent subscribers who are likely to churn in the future by taking the required retention policies to attract the likely churners and to retain them. Thereby, the potential loss of the company could be avoided. Instead of individualized client maintenance, the company can concentrate on a particular group of customers who are likely to churn. Because the companies can't bear to invest much time and cash for it. As the telecom clients are billions in number even a little part of stir prompts high loss of income.

Information mining strategies turn out to be a feasible alternative for predicting churn. Given a predefined forecast horizon, the goal is to predict the future churners over that horizon, given the data associated with each subscriber in the network. The input for this problem includes the data on past calls for each mobile subscriber, together with all personal and business information that is maintained by the service provider. In addition, for the training phase, labels are provided in the form of a list of churners. After the model is trained with the highest accuracy, the model must be able to predict the list of churners from the real dataset which does not include any churn label. In the perspective of the knowledge discovery process, this problem is categorized as predictive mining or predictive modeling.

II. RELATED WORK

Many kinds of research have been done in the field of CRM(Customer Relationship Management) in various industries for retaining customers and to develop strategies for building an efficient model so that specific group of customers can be aimed for retention. Decision trees, Regression models, Neural Networks, Clustering, Bayesian Models, SVM, etc are some of the famous data mining and statistical techniques used for churn prediction. In [1] the authors have proposed a hybrid learning model to predict churn in mobile Telecommunication networks. Their model is built using WEKA, a well-known tool of Machine Learning. They have proposed that DM (Data Mining) can detect the customers with a high propensity to churn, but not necessarily providing the reason of churn. The goal of their study was to show that hybrid models built on DM techniques can explain the churn behavior with more accuracy than single methods and that to some extent the reason of churn can be revealed. They used Logistic Regression in parallel with Voted Perceptron for classification and then combined with clustering for churn prediction.

There are many types of research on churn predictive model in the telecommunication. In [2] the researchers used a decision tree, neural network and logistic regression for customer classification and identified decision tree shows highest hit ratio among them. In [3] the author investigates the causes of telecom churn using Fuzzy Logic. In [4] author proposed an Artificial Neural Network (ANN) integrated prediction model for prepaid customers that could explain the reason of churn using the data set of complaints records from subscribers and thus the appropriate measure to be taken for retention strategy. In [5] the authors focused on Binomial logistic regression model for churn prediction and identified customer dissatisfaction, service usage, switching cost, and demographic variable affects customer churn.

In [6] the researchers applied Bayesian belief network to find out the most important factors that have effects on customer churn in the telecommunication industry and CAID algorithm is used to discretize continuous variable in churn. In [7] authors used Bayesian network classifiers for identifying the slope of the customer lifetime value (CLV) for long-life customers but used simple linear regression on the historical contributions of each customer to capture their individual life cycles. The slope was then separated into either positive or negative classes to represent increased or decreased spending. This variable was then used as the dependent variable in their study.

The decision trees are the most commonly used tool for classification & predictions of future events. The growth of such trees is completed in two major steps: building & pruning. During the first phase, the data set is partitioned recursively until most of the archives in each partition contain an equal value. The second phase then eliminates some branches which comprise the noisy data (those with the largest assessed error rate). CART, a Classification and Regression Tree, is created by the recursive division of an instance into subgroups until a definite standard has been met. The tree produces until the reduction of impurity falls below a user-defined threshold. All nodes in the decision tree are tested condition & the branching is based on the value of quality being tested. The tree is representing a group of multiple rule sets. When estimating a client dataset the arrangement is done by crossing through the tree until the leaf node is strained. The label of the leaf node (Churner / Non-Churner) is allocated to the client record under assessment.

In [8] the authors presented how to put into application grouping decision tree methods for churn examination in the telecommunication industry. An illustration set is used to carry out a test of customer churn issue using the ID3 decision tree. In their outcomes, they establish that the area of subscriber was the main classification characteristics that contributed to client churn, other than two minor reasons for a customer to churn. In research [9] they aimed at evolving a predictive model for client churn in pre-paid mobile telephony establishments. They applied decision trees methods like C5.0 with the neural network & it was exposed that based on improvement measure decision trees executed better than neural networks. An associated study was approved out [10] in which the researchers shared the J48 decision tree along the Genetic algorithm & constructed a hybrid evolutionary method for churn prediction in mobile networks. The author achieved 72% accurate consequences for the largest telecom company in evolving countries.

In [11] the author had the purpose of specifying the most appropriate model for churn prediction analysis. They showed an estimate of the different algorithms such as CART trees, neural networks& regression &confirmed their correctness in predicting customer churn. They originate that decision trees outperform rest of other methods with an overall correctness percentage of 82%. In [12] researchers proposed useful Naive Bayes, J48 & support vector machinery classifiers to process data so as to classify the important characteristics of the customers that help in forecasting churn of the bank clients. In their findings, they decided that achievement forecast of the loyal class is less than the prediction success rate % of the churn class. Additionally, they also originate that the J48 decision tree had enhanced performance related to other methods.

III. PROPOSED TECHNIQUE

The Churn Prediction System has three steps namely training, data preprocessing and testing.

i) Train dataset- First we have to train the dataset before we input raw data to the system for testing. Dataset has been taken from a telecom company operating in any country. This dataset contains many instances with several attributes; all extracted from customer services usage pattern like usage details of voice SMS and data. In the data preparation phase, data is collected, integrated and cleaned. Integration of data may require the extraction of data from multiple sources. Once the data has been arranged in tabular form, it needs to be fully characterized. Data needs to be cleaned by resolving any ambiguities, errors. Also redundant and problematic data items are to be removed at this stage. Not all fields of the database are always suitable for modeling purposes.

ii) Data preprocessing- The preprocessing phase has two steps:

1. Noise removal- In this step, noisy values have been removed to some extent; however, as we use FURIA for model building, thorough cleaning of data was not required at the pre-processing stage.

2. Feature selection- This is one of the most important and critical steps of implementation where only important and relevant features have been selected from large size dataset based on the domain knowledge: Decline in weekly spend rate, Decrease in outbound call count, Increase in days on zero balance, Balance burn out, Increase in calls to one competitor, Decrease in outbound call count, etc.

iii) Testing dataset- After the data is trained properly; we give raw data to the system for churn prediction. By using FURIA, the data will be tested and the model will predict in advance the possible churners in the future with more accuracy in less amount of time.

1. Fuzzy Unordered Rule Induction Algorithm (FURIA)

In this paper, we propose a novel fuzzy rule-based classification method called Fuzzy Unordered Rule Induction Algorithm, or FURIA which is a modification and extension of the state-of-the-art rule learner RIPPER. FURIA learns fuzzy rules instead of conventional rules and also unordered rulesets instead of rule lists in particular. This way, it becomes possible to model decision boundaries in a more flexible way. It makes use of an efficient rule stretching method to deal with uncovered examples. FURIA significantly outperforms the original RIPPER, as well as other classifiers such as C4.5, in terms of classification accuracy. Fuzzy rules are more general than conventional rules and it has a number of advantages. For example, conventional (non-fuzzy) rules produce models with "sharp" decision boundaries and similarly, abrupt transitions between different classes. This property is questionable and not very intuitive. Instead, one would expect the support for a class provided by a rule to decrease from "full" (inside the core of the rule) to "zero" (near the boundary) in a gradual rather than an abrupt way. One of the main characteristics of a fuzzy rule is that it has "soft" boundaries. Admittedly, if a definite classification decision has to be made; soft boundaries should be again turned into crisp boundaries. However, in the fuzzy case, these boundaries are potentially more flexible. For example, by using suitable aggregation operators for combining fuzzy rules, they are not necessarily axis-parallel.

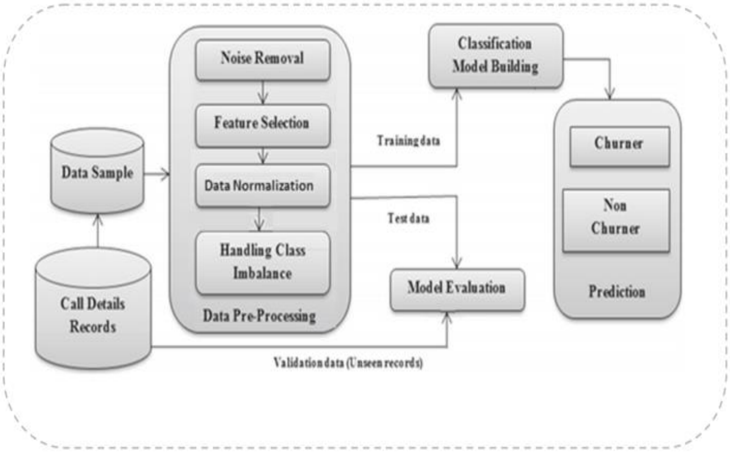

Fig 1: Fuzzy Model of churn prediction

In order to build a model, we have used different classifiers in our methodology. In recent years, fuzzy is getting more attention to classification in telecom churn prediction due to the nature of the data. In churn prediction domain, data is very noisy and fuzzy classifiers especially Vaguely Quantified Nearest Neighbors (VQNN) effectively handles such data. In the next subsection, a short description of only fuzzy classifiers with promising results is given. In order to classify an object based on likeness and similarity with another one, the Fuzzy K nearest neighbor (FNN) algorithm is introduced. Let's suppose an object y belongs to class C which can be written as:

$$C'(y) = \sum_{x \in N} R(x, y)\, C(x)$$

............Equation (1)

Where N is the y's K nearest neighbors, and R(x, y) is the [0, 1]—similarity of x and y. It can be written as follows.

$$R(x, y) = \frac{||y - x|| - 2/(m-1)}{\sum_{j \in N} ||y - j|| - 2/(m-1)}$$

...........Equation (2)

Where, R denotes the Euclidean norm, and 'm' is a parameter which controls the weighting of the similarity, in our case 'm' is set to the default value 2. Assuming crisp classes, Algorithm 1 shows an application of the FNN algorithm that classifies a test object 'y' to the class with the highest resulting membership. The idea behind this algorithm is that the degree of closeness of neighbors should influence the impact that their class membership has on deriving the class membership for the test object.

1.1. Algorithm 1

The fuzzy algorithm fuzzy(U,C,y,K).
U, the training data;
C, the set of decision classes;
y, the object to be classified;

K, the number of nearest neighbors.
(1) N ← getNearestNeighbours (y, K);
(2) ∀C ∈ C
(3) C(y) = P x∈N R(x, y) C(x)
(4) outputarg max C ∈ C (C(y))

2. C4.5 Algorithm

The C4.5 algorithm uses the 'divide and conquer' method to construct a model based on a tree structure. Nodes in the tree represent features, with branches representing possible values connecting the features. A leaf representing the class terminates a series of nodes and branches. Initially, the method starts to search an attribute with best information gain at the root node and divide the tree into sub-trees. Similarly, each sub-tree is further separated recursively following the same rule. The partitioning stops if the leaf node is reached or there is no information gain. Once the tree is created, rules can be obtained by traversing each branch of the tree.

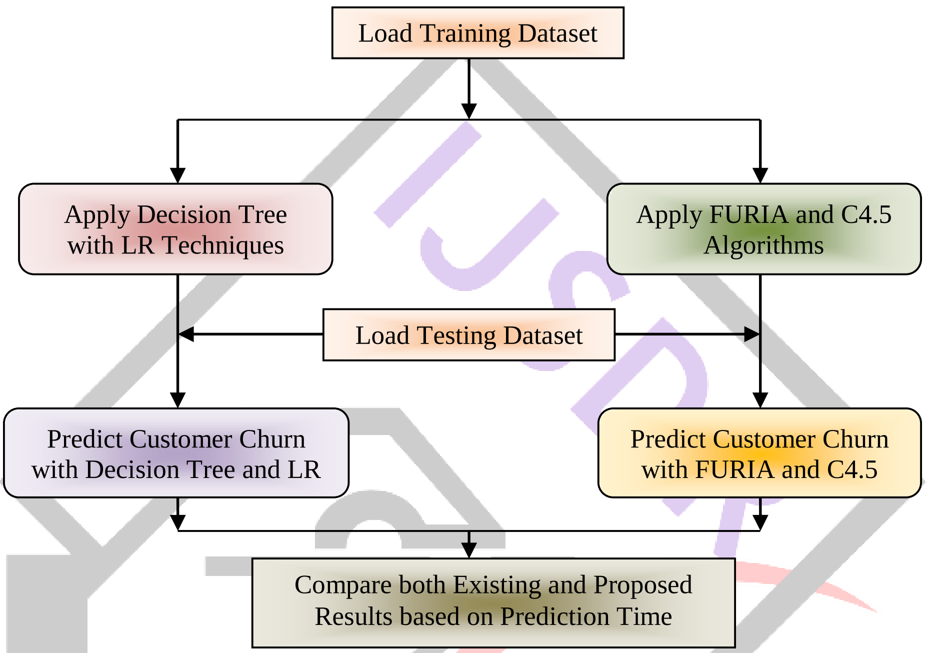

Fig 2: System Architecture

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The following graph shows the accuracy, execution time, recall and precision of FURIA compared to the Decision tree. The True positive and false positive rate of Fuzzy classifiers, C4.5 is higher as compared to other classifiers. The maximum churners have been captured by Fuzzy Classifiers as TP rate is very high.

A. Accuracy:

The following graph shows the accuracy rate of FURIA in predicting churn compared to the Decision tree.

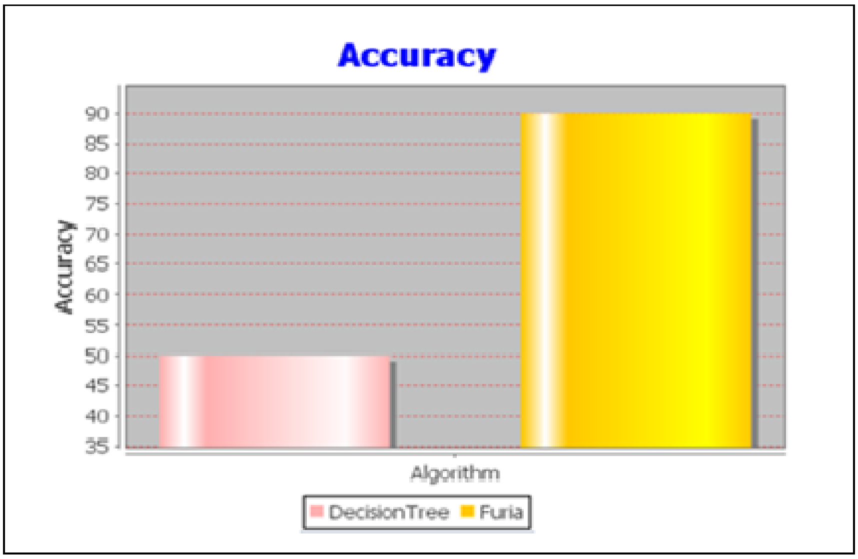

Fig. A: Accuracy graph of FURIA and Decision tree

The confusion matrix is a standard table used to measure the accuracy of both churner and non-churner classes. As shown in the graph the accuracy of FURIA is high compared to the Decision tree. So FURIA gives more accurate results while predicting churn.

B. Execution Time:

The following graph shows the execution time of FURIA and Decision tree of the given dataset in predicting customer churn.

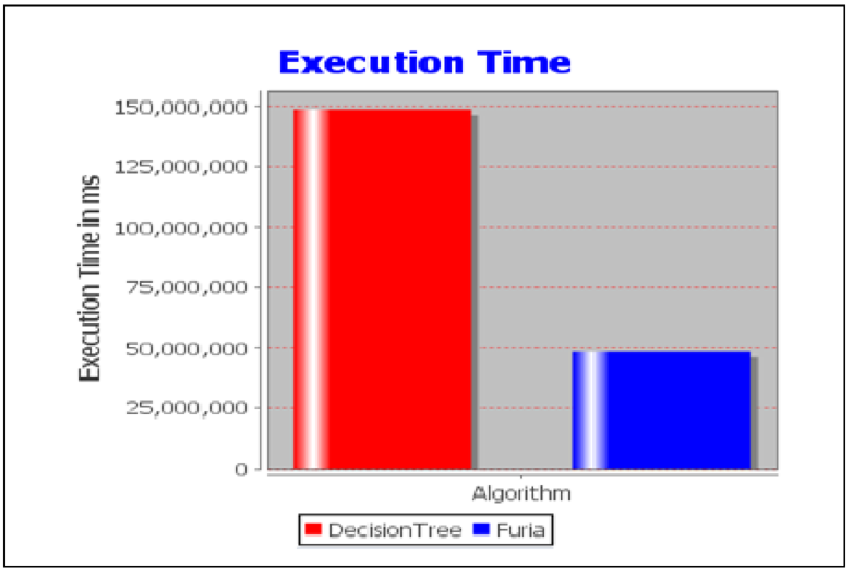

Fig. B: Execution time graph of FURIA and Decision tree

A good prediction model is the one which gives more accurate results in less amount of time. Only small dataset input can be processed by the existing system. As the customers' databases are billions in number, it takes a lot of time for execution and to predict the churn in advance. But in FURIA large dataset input can be given and takes less time for predicting the churn with more accuracy.

C. Precision:

The following graph shows the precision value of FURIA when compared to Decision tree in churn prediction.

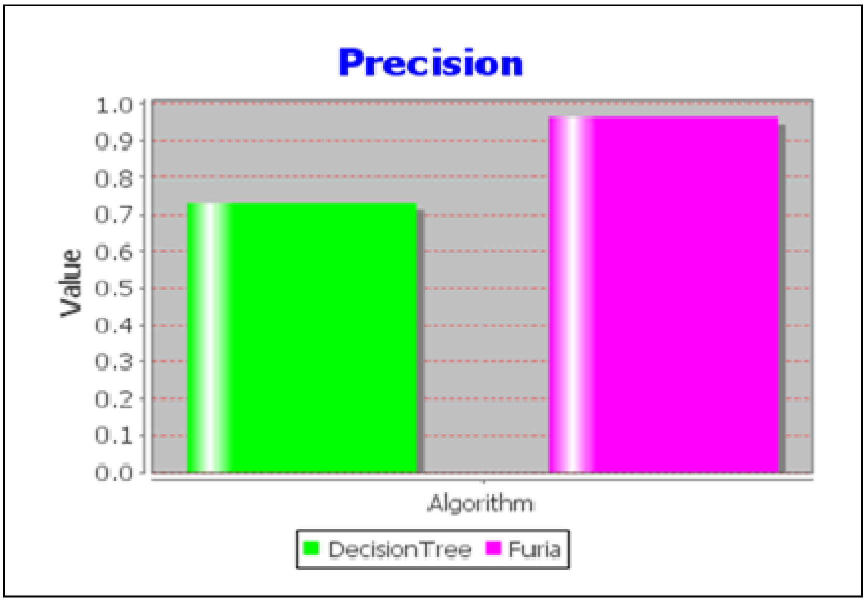

Fig. C: Precision graph of FURIA and Decision tree

Precision and recall are the measures that actually evaluate the accuracy of the churn prediction model. Precision is basically the ratio of relevant records (TP) identified to the total number of irrelevant and relevant records. It is usually expressed as a percentage. In our example, precision will be calculated as TP/ (TP + TN). From the graph, it is clear that the precision value of FURIA is high when compared to the Decision tree.

D. Recall:

The following graph shows the recall value of FURIA comparing to the Decision tree.

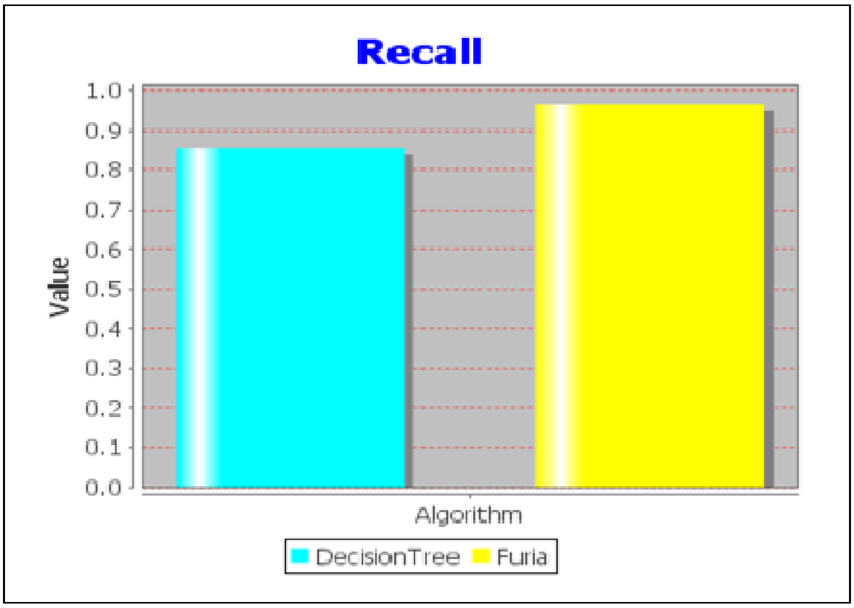

Fig. D: Recall graph of FURIA and Decision tree

The recall is the ratio of the number of relevant records (TP) retrieved to the total number of relevant records in the dataset. It is also expressed in percentage as Recall i.e. TP/ (TP + FN). The recall value of FURIA is high when compared to the Decision tree in churn prediction.

V. CONCLUSION

In the proposed system, Fuzzy Unordered Rule Induction Algorithm is used for predicting customer churn in advance. FURIA learns fuzzy rules instead of conventional rules and unordered rulesets instead of rule lists. Fuzzy rules have soft boundaries, which is one of their main characteristics. And it makes use of an efficient rule stretching method. FURIA provides more accurate results in predicting the customers who are more likely to churn in future than Decision tree. The maximum churners have been captured by Fuzzy Classifiers as TP rate is very high. Using this model the telecom companies can predict in advance which customers are at risk of leaving, and can target on those customers. So the companies can design retention policies to maintain those customers. This can help the companies in saving a lot of revenues. When compared to a Decision tree, the FURIA can execute a large dataset input in very less amount of time with more accurate results.

REFERENCES

[1] A Hybrid Churn Prediction Model in Mobile Telecommunication Industry, Georges D. Olle Olle and ShuqinCai, International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 4, No. 1, February 2014.

[2] Beomsoo Shim, Keunho Choi, YongmooSuh, ―CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns‖, Expert Systems with Applications 39 (2012) 7736–7742.

[3] K. H. Liao and H. E. Chueh, "Applying fuzzy data mining to telecom churn management," Intelligent Computing and Information Science, 2011, pp. 259-264.

[4] J. O. Daramola, O. O Oladipupo, and G. A. Musa, "A data mining process framework for churn management in the mobile telecommunication industry," IJCSS, vol. 1. no. 1, 2009, pp. 34-40.

[5] Abbas Keramati, Seyed M.S. Ardabili, ―Churn analysis for an Iranian mobile operator‖, Telecommunications Policy 35 (2011) 344–356.

[6] PýnarKisioglu, Y. IlkerTopcu, ―Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey‖, Expert Systems with Applications 38 (2011) 7151–7157.

[7] BAESENS, B., VERSTRAETEN, G., VAN DEN POEL, D., EGMONT-PETERSON, M., VAN KENHOVE, P. & VANTHIENEN, J. (2004) Bayesian Network Classifiers for Identifying the Slope of the Customer Lifecycle of Long-Life Customers. European Journal of Operational Research, 156, 508-523.

[8] K. Binti Oseman, S. Binti M. Shukor, N. Abu Haris, and F. bin Abu Bakar, ―Data Mining in Churn Analysis Model for Telecommunication Industry,‖ J. Stat. Model. Anal., vol. 1, no. 19–27, 2010.

[9] A. T. Jahromi, M. Moeini, I. Akbari, and A. Akbarzadeh, ―A Dual-Step Multi-Algorithm Approach For Churn Prediction in Pre-Paid Telecommunications Service Providers,‖ RISUS. J. Innov. Sustain, vol. 1, no. 2, 2010.

[10] V. Yeshwanth, V. V. Raj, and M. Saravanan, ―Evolutionary Churn Prediction in Mobile Networks Using Hybrid Learning,‖ in Twenty-Fourth International FLAIRS Conference, 2011, pp. 471–476.

[11] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, ―Churn prediction: does technology matter,‖ Int. J. Intell. Technol., vol. 1, no. 2, 2006.

[12] M. Kaur, K. Singh, and N. Sharma, ―Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers,‖ Int. J. Recent Innov.Trends Comput.Commun.vol. 1, no. 9, pp. 720–725, 2013.