Customer Retention in Banking Sector using Predictive Data Mining Technique

K. Chitra*, B.Subashini[#]

*Dept of Computer Applications
Thiagarajar School of Management
Madurai, Tamil Nadu, India
[#]Dept of E-Commerce
VVV College for Women
Virudhunagar, Tamil Nadu, India
*chitra@tsm.ac.in
#subamca2003@yahoo.com

Abstract- **Churn in the banking sector is a major problem today. Losing the customers can be very expensive as it costs to acquire a new customer. In this paper, we have made a solution for the churn problem in banking sector using data mining technique. Predictive data mining techniques are useful to convert the meaningful data into knowledge. As predicting churn is more important for a bank, we have used Classification and Regression Trees to yield a better overall classification rate.**

Keywords- **Predictive Data Mining, Churn Prediction, Classification models, Banking sector**

I. INTRODUCTION

Customer Retention is an increasingly pressing issue in today's ever-competitive commercial arena. More competition and increased regulation made it more difficult for banks to stand out from the crowd. Studies done at international and national levels indicate the significance of a few imperatives for commercial banks for their survival and growth. These are

1. Customer retention
2. Focus on technology
3. Focus on specific market segments
4. Enhanced productivity and efficiency

Above four factors, Customer retention is first and effective method for the growth of the banks. Customer Relationship Management tools have been developed and applied in order to improve customer acquisition and retention and to support important analytical tasks such as predictive modeling and classification [10]. Typically CRM applications hold a huge set of information regarding each individual customer. This information is gained from a customers' activity at the bank. Database entries are scored using a statistical model defined over various attributes which characterize the customers. These attributes are often called predictor variables. This paper tries to propose a solution for Churn Problem using Classification and Regression Tree.

II. CHURN ANALYSIS

Customer churn is the term used in the banking sector tries to denote the movement of customers from one bank to another. In the banking industry, identifying probable churn customers has increased in its importance in the recent past [8]. In banking domain, we define a churn customer as one who closes all his/her accounts and stops doing business with the bank. There are many reasons for a customer to close the account(s). For example, a person creates an account for a specific purpose and closes it immediately after the purpose is solved. Or a person is relocated and has to move to another place and hence closes all the accounts. Or a customer may stop transacting with the bank just because of the unavailability of bank's ATMs in important places and hence close his/her accounts. The problem here is that, in real world scenario, the bank does not always capture this kind of feedback data. Hence, no further analysis can be done and this type of churning behaviors could not be stopped. This leaves us in a situation where we need to think which kind of churn patterns are possible to identify.

A. Churn Prediction

For finding answers to the questions who and why is likely to churn a classification of the customers is needed. Churn prediction deals, therefore, with the identification of customers likely to churn in the near future. The basis for this is historical data, containing information about past churners. A comparison is made between these churners and existing customers. As likely churners are identified customers for which the classification suggests similarity to prior churners. In this paper, we explain the kind of raw data available in real-time banking scenario, provide a guideline to convert raw data into meaningful data and finally convert the meaningful data into knowledge using predictive data mining techniques. For churn prediction to be truly meaningful, the root causes of historic churn must be thoroughly understood. Measuring the accuracy of the churn prediction is another important step addressed by the solution.

III. PREDICTIVE DATA MINING

Data mining is the process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules [2]. It can also be defined as the process of selecting,

exploring and modeling large amounts of data to uncover previously unknown data patterns for business advantage.

The classification of large data sets is an important problem in data mining. The classification problem can be simply stated as follows. For a database with a number of records and for a set of classes such that each record belongs to one of the given classes, the problem of classification is to decide the class to which a given record belongs. But there is much more to this than just simple classifying.

Using data mining, descriptive models and predictive models can be built [5]. Descriptive models are built on the concept of unsupervised learning and predictive models are built on the concept of supervised learning [1]. In a predictive model, one of the variables (the target variable or response variable) is expressed as a function of the other variable. In a predictive model, one of the variables (the target variable or response variable) is expressed as a function of the other variables. In the churn prediction problem, the response variable, i.e., the future status of the customers can take only two values viz. Active or Churn. Therefore predictive classification techniques are used for churn modeling.

There are many predictive classification techniques namely nearest neighbor technique, decision tree technique, linear discriminant technique, naïve bayes technique, etc. In this paper, decision tree technique is used. The specific requirements that should be taken into consideration while designing any decision tree construction algorithms for data mining are that

    a. The method should be efficient in order to handle a very large-sized databases
    b. The method should be able to handle categorical attributes.

B. Decision Tree

A decision tree is a classification scheme which generates a tree and a set of rules, representing the model of different classes, from a given data set. The set of records available for developing classification methods is generally divided into two disjoint subsets- a training set and a test set. The former is user for deriving the classifier, while the latter is used to measure the accuracy of the classifier. The accuracy of the classifier is determined by the percentage of the test example that is correctly classified.

We categorize the attributes of the records into two different types. Attributes whose domain is numerical are called the numerical attributes, and the attributes whose domain is not numerical are called the categorical attributes. There is one distinguished attribute called the class label. The goal of the classification is to build a concise model that can be used to predict the class of the records whose class label is not known.

C. Decision tree Construction algorithm

A number of algorithms for inducing decision trees have been proposed over the years. However, they differ among themselves in the methods employed for selecting splitting attributes and splitting conditions. The first type of algorithm is the classical algorithms which handle only memory resident data. Efficiency and scalability are the fundamental issues concerning the classification of data ([4], [7]). The second category of algorithms can handle the efficiency and scalability issues. These algorithms remove the memory restrictions and are fast and scalable.

D. CART

CART is one of the popular methods of building decision trees in the machine learning community [3]. CART builds a binary decision tree by splitting the records at each node, according to a function of single attribute. CART uses the gini index for determining the best split. CART follows the above principle of constructing the decision tree. Outline of this method is for the sake of completeness. The initial split produce two nodes, each of which we new attempt to split in the same manner as the root node. Once again, we examine all the input fields to find the candidate splitters. If no split can be found that significantly decreases the diversity of a given node, we label it as a leaf node. Eventually only leaf nodes remain and we have grown the full decision tree. The full tree may generally not be the tree that does the best job of classifying a new set of records, because of over fitting.

At the end of the tree-growing process, every record of the training set has been assigned to some leaf of the full decision tree. Each leaf can now be assigned a class and an error rate. The error rate of a leaf node is the percentage of incorrect classification at that node. The error rate of an entire decision tree is a weighted sum of the error raters of all the leaves. Each leafs contribution to the total is the error rate at that leaf multiplied by the probability that a record will end up in there.

The information equation is $-\sum p * \lg(p)$

Where p is probability value occurring in a particular node of the tree ([6], [9]).

E. Data Set

Usage data is the key to investigating previous churn patterns and enabling the prediction of potential churners. There are four sets of data variables: customer behavior, customer perceptions, customer demographics and macro environment variables.

Customer behavior identifies which parts of the service a customer is using and how often is he using them.

Customer perceptions are defined as the way a customer apprehends the service. They can be measured with customer surveys and include data like overall satisfaction, quality of service, problem experience, satisfaction with problem handling, pricing, etc.

Customer demographics are some of the most used variables for churn prediction. They include age, gender, level of education, social status, geographical data, etc.

Macro-environment variables identify changes in the world, different experiences of customers, which can affect the way they use a service.

The size of gathered data is usually very large, which results in high dimensionality, making the analyze a complex and challenging task. Therefore, before beginning to use a churn prediction method a data reduction technique is used, deciding with application domain knowledge which attributes can be of use and which can be ignored. Missing values should also be regarded - on attribute level these can be ignored if they are with low significance, whereas on record level they have to be replaced with a reasonable estimate, for example using interpolation. Providing a good estimate for this missing values is an important issue for proper churn prediction.

TABLE I. TABLES USED IN DATA SET

Table name	Attribute	No. of records
Customer	no, name, dob, age, gender, qualification	10,000
Account	ano, type, balance, flag	10,000
Transaction	ttype, date, ano, amount	55,000

F. Data Preparation

If we fix the timeline the quality of the data becomes less. In this paper, we suggest to consider a dynamic timeline, which varies for each customer. We adapt a dynamic timeline for each customer and hence avoid the problem of not training the model properly. This concept applies to churn records. Because churning time varies from customer to another customer. For active records, we can consider the behavior in any period. In our analysis, we consider the behavior of last 3 months before last transaction date of the active customers. The number of months of data to be considered for churn analysis is a business problem. Generally, considering the transaction activities of 3 months would suffice the requirement.

For example the customer table contains customer details like Customer name, date-of-birth, etc. Master table contains Account Number and Balance amount. Based on dynamic timeline, we can get the no. of records from these tables such as Master Table and Customer Table.

Suppose if we take the dynamic timeline as 3 months, the total no. of records of each table are compressed as follows.

TABLE II. NO. OF RECORDS IN EACH TABLE

Table name	No. of records
Customer	1000
Account	1000
Transaction	6000

The next thing, we are more concentrating on the splitting attribute. Splitting attribute, with every node of the decision tree, there is an associated attribute whose values determine the partitioning of the data set when the node is expanded. That is, under what factor the decision tree grows more.

The qualifying condition on the splitting attribute for data set splitting at a node is called the splitting criterion at that node. For a numerical attribute, the criterion can be an equal or an inequality. For a categorical attribute, it is a membership condition on a subset of values. If we have a splitting attribute as age, we can set the splitting criterion as age >50 or <50. Otherwise if we have a splitting attribute as gender, we can set the splitting criterion as female or male.

The target variable is churn status that is active or churn. The dataset is split into training dataset and testing dataset in building the classification trees.

G. Churn model using CART

The process of learning and rules abstraction is identified as churn modeling. Classification tree is built using CART on the training dataset using the following specifications: CART uses no stopping rule that could be relied on to discover optimal tree. So the tree is over grown and then pruned back which ensures that important patterns are not overlooked by stopping too soon. CART does binary splitting that are more sparing with data and detect more patterns before too few data are left for learning. It produces rules for the target variable as a function of other fields in the dataset that are previously identified as derived variables. Gini concentration coefficient is used to summarize power curves of prediction.

Next we have to find out the success rate for training data set and test data set.

Let us assume that we have totally 1000 records. Out of 1000 records 700 records are for training data and 300 records are for test data. Predicted class data is calculated from the decision rules generated by CART.

TABLE III. SUCCESS RATE FOR THE TRAINING DATA

Class	Tot. Samples	Predicted Active	Predicted Churn	Success%
Active	545	455	90	83.48
Churn	155	15	140	90.32

TABLE IV. SUCCESS RATE FOR THE TEST DATA

Class	Tot. Samples	Predicted Active	Predicted Churn	Success %
Active	216	179	37	82.87
Churn	84	7	77	91.66

The success rate of active customers is relatively less. This is because, some of the customers, although their status is marked as active, exhibited churn characteristics.

IV. CONCLUSIONS

The solution exploits the power of operational data to perform churn analysis, identify those customers most likely to churn and trigger retention campaign strategies. By using data mining statistical algorithms, the solution discovers the churn patterns within past churners' behavior and uses this knowledge to assign a potential churn score rating to current customers. The study thus predicts the future churn for banking customers and can then be extended, thereby helping formulate 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, bank managers can develop a customized approach to customer retention activities within the context of their Customer Relationship Management efforts.

REFERENCES

[1] Margaret Dunham "Data Mining: Introductory and Advanced Topics"

[2] Pujari Arun. K. "Data Mining Techniques"

[3] Jiawei Han "Data Mining: Concepts and Techniques"

[4] Stephane Tuffery "Data Mining et Statistique Decisionnelle"

[5] Oded Maimon, Lior Rocach "The Data Mining and Knowledge Discovery Handbook"

[6] Daniel T. Larose "Data Mining Methods and Models"

[7] Pang-Ning Tan, Michael Steinbach, Vipin Kumar "Introduction to Data Mining"

[8] Bhasin, M.L. "Data Mining: A Competitive Tool in the Banking and Retail Industries"

[9] Cowley, S., "Data Mining"

[10] Berry, M. J. A. and Linoff G. S. "Mastering Data Mining The art of Customer Relationship Management"