Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers

Manjit Kaur
Department of Computer Scienc
Punjabi University
Patiala, India
manjit8718@gmail.com

Dr. Kawaljeet Singh
University Computer Centre
Punjabi University
Patiala, India
singhkawaljeet@rediff.com

Dr. Neeraj Sharma
Department of Computer Scienc
Punjabi University
Patiala, India
sharmaneeraj@hotmail.com

***Abstract*— The socio economic growth of the country is mainly dependent on the services sector. The financial sector is one of these services sector. Data mining is evolving into a strategically important dimension for many business organizations including banking sector. The churn problem in banking sector can be resolved using data mining techniques. The customer churn is a common measure of lost customers. By minimizing customer churn a company can maximize its profits. Companies have recognized that existing customers are most valuable assets. Customer relationship management (CRM) can be defined as the process of acquiring, retaining and growing profitable customer which requires a clear focus on service attributes that represent value to the customer and creates loyalty. Customer retention is critical for a good marketing and a customer relationship management strategy. The prevention of customer churn through customer retention is a core issue of Customer relationship management. Predictive data mining techniques are useful to convert the meaningful data into knowledge. In this analysis the data has been analyzed using probabilistic data mining algorithm Naive Bayes, the decision trees algorithm (J48) and the support vector machines(SMO).**

Index Terms — *customer churn; banking sector; predictive data mining; CRM*

*****

I. INTRODUCTION

Classification of the services sector by Central Statistical Organization (CSO) consists of four broad categories, first is “trade, hotels and restaurants”, second is “transport, storage and communication”, third is “financing, insurance, real estate and business services” and the fourth is “community, social and personal services” [10]. In Yojna, September 2011 issue, the services sector has been highlighted as the lifeline for the *socio economic growth* of a country. It is today the largest and fastest growing sector globally contributing more to the global output and employing more people than any other sector. Financial services sector is one of these sectors and data mining assists many kinds of analysis work in this area such as:

- *Financial product cross-selling*
- *Market segment analysis*
- *Fraud detection*
- *Customer churn analysis*

Customer Retention is an increasingly pressing issue in today's ever-competitive commercial arena. Churn is defined as the propensity of a customer to stop doing business with an organization and subsequently moving to some other company in a given time period. The customer churn is a common measure of lost customers. Customer retention rate has a strong impact on the customer lifetime value, and understanding the true value of a possible customer churn will help the company in its customer relationship management (CRM).

The subject of customer retention, loyalty, and churn is receiving attention in many industries. This is important in the customer lifetime value context. A company will have a sense of how much is really being lost because of the customer churn and the scale of the efforts that would be appropriate for retention campaign. The mass marketing approach cannot succeed in the diversity of consumer business today. Customer value analysis along with customer churn predictions will help marketing programs target more specific groups of customers.

Personal retail banking sector is characterized by customers who stays with a company very long time. Customers usually give their financial business to one company and they won’t switch the provider of their financial help very often. In the company’s perspective this produces a stable environment for the customer relationship management. Although the continuous relationships with the customers the potential loss of revenue because of customer churn in this case can be huge.

Data mining can be used to maintain customer relationship management. Various techniques are available to analyze and infer customer behavior in future using predictive data mining.

II. CHURN ANALYSIS

In the banking sector, the term churn denotes the movement of customers from one bank to another. In the domain of banking, churn customer is one who closes all his/her accounts and stops doing business with the bank and the reasons for a customer to close the account are many. One may creates an account for a specific purpose and closes it after the purpose is solved. A person may be relocated to another place and thus closes all the accounts. The problem is that, in real world this kind of feedback data is not always captured by the bank. Thus further analysis cannot be done and this type of churning behaviors left unrevealed. Thus we need to think which kind of churn patterns are possible to identify.

A. *Churn Prediction*

Personal retail banking sector is a market sector where a customer does not regularly switching from one company to another. Customers usually give their banking business to one or two banks for long periods of time. Identifying the churn before hand and taking necessary steps to retain the customers would increase the overall profitability of the organization. Losing customers not only leads to opportunity lost because of reduced sales, but also to an increased need for attracting new customers, which is five to six times more expensive than customer retention. Banks have the source of customer data in the form of daily transactions and operations. Banks generate a large amount of data through operations such as credit card processing, ATM usage, cash withdrawals and deposits and much more. This data can provide valuable information about customers' behavior toward the bank presently and in near future and help in classifying the customers. With the use of data mining, this customer data can be organized and extracted to facilitate future predictions about the customer retention and management decisions which can be a boon to the financial company.

III. CHURN PREDICTIVE MODEL

Data Mining was used to predict whether a particular customer churned over a given period of time. Data mining is the process of exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. 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. In a predictive model, one of the variables (the target variable or response variable) is expressed as a function of the other variable. In the churn prediction problem, the response variable, i.e., the future status of the customers can take only two values viz. Active/Loyal or Churn. Therefore predictive classification techniques are used for churn modeling. There are many predictive classification techniques namely nearest neighbor technique, decision tree technique,, naive bayes technique, etc. In this research, naive bayes, decision trees and support vector machine technique are used. The following are the various classification techniques:

A. Statistical- Based Algorithms

a. Regression: Simple regression analysis is a statistical tool That gi ves us the ability to estimate their mathematical relationship betw een a dependent variable (usually called y) and an independent va riable (usually called x).The dependent variable is the variable for which we want to make a prediction.

b. Bayesian Classification: It is based on Bayes rule of conditional p robability. By analyzing the contribution of each independent att ribute a conditional probability is determined. A classification is made by combining the impact that different attributes have on th e prediction to be made.

c. K-nearest Neighbors: It is based on the use of distance measure. When a classification is to be made for a new item its distance to each item in the training set must be determined. Only the k clos est entries in the training set are considered further.

B. Decision Tree-Based Algorithms

a. ID3: This technique is based on information theory and attempts t o minimize the expected number of comparisons. The basic idea of induction algorithm is to ask questions whose answers provide the most information. The objective is to partition the given data set into subsets where all elements in each final subset belong to t he same class

b. C4.5 and C5.0: At each node of the tree, C4.5 chooses one attribu te of the data that most effectively splits its set of samples into su bsets enriched in one class or the other. Its criterion is the normali zed information gain (difference in entropy) that results from cho osing an attribute for splitting the data. The attribute with the hig hest normalized information gain is chosen to make the decision. The C4.5 algorithm then recourses on the smaller subsists. C5.0 i s a commercial version of C4.5

c. CART: Classification and Regression trees is a technique that ge nerates a binary decision tree CART uses no stopping rule that co uld be relied on to discover optimal tree. So the tree is over grow n 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 t oo few data are left for learning.

C. Neural Network Based Algorithm

In more practical terms neural networks are non-linear statistical data modeling tools. They can be used to model complex relation ships between inputs and outputs or to find patterns in data. In ne ural network a model representing how to classify any given data base tuple is constructed. When a tuple must be classified, certai n attribute values from that tuple are input into the directed graph at the corresponding source node. The output value that is generat ed indicates the probability of membership. The learning process modifies the labels in the graph to better classify tuples.

IV. METHODOLOGY

Our approach consists of data sampling, data preprocessing, model construction, and model evaluation phases. Data sampling selects a set of customers with the required information. The data preprocessing phase includes data cleaning. Data cleaning removes the irrelevant information which includes wrong spelling words caused by human errors, special mathematical symbols, missing values, duplicated information, and so on.

In the model construction phase, we build a classification/prediction model that predicts the potential behavior of customers in the near future.

A. Data sampling :

The database had already a churn variable for current churned customers and loyal for active customers. The dataset contained 2000 records of both churned and active customers that included. The data was divided into 70% training and 30% validation.

Table name	**Attribute**
Customer	accno, name, dob, age, gender
Account	ano, type, balance,
Transaction	ttype, date, tnum

B. Data preprocessing

In this work, we focused on data cleaning in this phase.

a. Data cleaning

Noise is the irrelevant information which would cause problems for the subsequent processing steps. Therefore, noisy data should be removed. This irrelevant information includes special symbols like mathematical symbols and punctuation marks, missing values, duplicated information etc. This noise can be removed by finding their locations and using the correct values to replace them, or some times by deleting them if the missing values are too many.

Customer profiles: describe the demographic grouping of customers basing on their common characteristics that include age, gender, level of education, etc.

Transactional details: These include branch transactions, ATM transactions, deposit, withdrawal, internet transactions etc.

C. Model construction

a. Naïve Bayes

		Predicted class	
		Active	Churn
Actual class	Active	TN	FP
	Churn	FN	TP

Naive Bayes learning generates a probabilistic model of the observed data. Despite its simplicity, Naive Bayes has been verified to be competitive with more complex algorithm such as neural network or decision tree in some domains. Given a training set of instances, each is represented as a vector of features [x1,x2,…..,xd], the task is learning from the data to be able to predict the most probable class of a new instance whose class is unknown. Naïve Bayes employs the Bayes's theorem to estimate the probabilities of the classes.

$$P(y_j| x_1,x_2,\ldots,x_d) = \frac{P(y_j)\, P(x_1,x_2,\ldots,x_d \mid y_j)}{P(x_1,x_2,\ldots,x_d)}$$

Where $P(y_j)$ is the prior probability of class which is estimated as its occurrence frequency in the training data. $P(y_j| x_1,x_2,\ldots,x_d)$ is the posterior probability of class y_j after observing the data. $P(x_1,x_2,\ldots,x_d \mid y_j)$ denotes the conditional probability of observing an instance with the feature vector $[x_1,x_2,\ldots,x_d]$, among those having class y_j. And $P(x_1,x_2,\ldots,x_d)$ is the probability of observing an instance with the feature vector $P(x_1,x_2,\ldots,x_d)$ regardless of the class. Since the sum of the posterior probabilities over all classes is one, the denominator on equation's right hand side is a normalizing factor and can be omitted.

$$P(y_j| x_1,x_2,\ldots,x_d) = P(y_j)\, P(x_1,x_2,\ldots,x_d \mid y_j)$$

An instance will be labeled as the particular class which has the highest posterior probability.

b. Decision Trees

The C4.5 technique is one of the decision tree families that can produce both decision tree and rule-sets; and construct a tree. Besides that, C4.5 models are easy to understand as the rules that are derived from the technique have a very straightforward interpretation. J48 classifier is among the most popular and powerful decision tree classifiers. C5.0 and J48 are the improved versions of C4.5 algorithms. WEKA toolkit package has its own version known as J48. J48 is an open source implementation of C4.5 algorithm in weka. J48 adopts a greedy approach in which decision tree is constructed in a top down recursive divide and conquer manner. The decision tree algorithm works as follows:

Create a node N;

if tuples in D are all of the same class, C then

return N as a leaf node labeled with the class C;

if attribute list is empty then

return N as a leaf node labeled with the majority class in D; // majority voting

apply Attribute selection method(D, attribute list) to find the "best" splitting criterion;

label node N with splitting criterion;

if splitting attribute is discrete-valued and multiway splits allowed then // not restricted to binary trees

attribute list attribute list _ splitting attribute; // remove splitting attribute

for each outcome j of splitting criterion // partition the tuples and grow subtrees for each partition

let Dj be the set of data tuples in D satisfying outcome j; // a partition

if Dj is empty then

attach a leaf labeled with the majority class in D to node N;

else attach the node returned by Generate decision tree(Dj, attribute list) to node N; endfor

return N;.

c. Support Vector Machine

In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other.

SMO which is an open source implementation of support vector machines in Weka.

D. Evaluation Criteria

For classification, the accuracy estimate is the overall number of correct classifications divided by the total number of tuples in the initial data. This provides a good indication of how well the classifier will perform on unseen data. A confusion matrix

Definitions:

- *True positive (TP):* Number of positive cases correctly predicted.
- *False negative (FN):* Number of positive cases wrongly predicted as negative.
- *False positive (FP):* Number of negative cases wrongly predicted as positive
- *True negative (TN):* Number of negative cases correctly predicted.

From the confusion matrix the following measures, among others, can be obtained.

Accuracy: The percentage of correctly classified instances over the total number of instances.

True positive rate (TPR) or sensitivity: fraction of positive instances predicted correctly.

False positive rate (FPR): fraction of negative instances wrongly predicted as positive.

Precision: fraction of records that actually turn out to be positive in the group the classifier has declared as positive. The higher the precision is, the lower the number of false positive errors committed by the classifier.

Recall: Fraction of positive instances correctly predicted by the classifier. Its value is equivalent to true positive rate. The higher the value of recall the fewer the number of instances misclassified as negative.

E. RESULT AND ANALYSIS

In this analysis, we have experimented with probabilistic classification technique namely Naïve bayes, decision trees (J48) and support vector machine(SMO). Total instances in the dataset were split into 70 % training set and the remaining 30% test data. We used naïve bayes to trace out significant customer characteristics to predict churn. The results are evaluated using confusion matrix:

True class	Total samples	Predicted churn	Predicted Loyal	Success %
Churn	58	58	0	100
Loyal	506	33	473	93.05

Table 1. CONFUSION MATRIX SUCCESS RATE(NAÏVE BATES)

The second technique used is decision tree algorithm J48. It yielded the following results:

True class	Total samples	Predicted churn	Predicted Loyal	Success %
Churn	54	54	0	100
Loyal	510	6	504	98.82

Table 2. CONFUSION MATRIX SUCCESS RATE(DECISION TREES)

The third technique used is support vector machine algorithm SMO and the results are :

True class	Total samples	Predicted churn	Predicted Loyal	Success %
Churn	58	58	0	100
Loyal	506	25	481	95.05

Table 3. CONFUSION MATRIX SUCCESS RATE(SMO)

However, the prediction success rate of Churn class is more than the prediction success rate of Loyal class. The accuracy of the decision trees is higher than the NB classifier and the SMO. However the predictive accuracy of NB classifier is comparable to other classification techniques. Thus, this model with higher prediction success rate of Churn class has to be chosen for reaping higher benefits.

V. CONCLUSION

Data mining aims to extract knowledge and insight through the analysis of large amounts of data using sophisticated modeling techniques. Thus, it is essential that the customers, indicated by the churn model, to become churn should be focused. If the churn prevention program is effective, the bank can look forward to reaping significant benefits from its efforts. In this paper, we have given a detailed guideline of converting raw customer data of a bank into useful data and then convert this data info useful information using data mining techniques. We have extracted the data for chosen attributes from raw customer data for a chosen set of 2000 customers. We used naive bayes, decision trees and support vector machine classifier to recognize significant customer characteristics to predict churn However, the prediction success rate of Churn class is more than the prediction success rate of Loyal class. This research work predicts the future churn of bank customers which can be addressed, by intervention so that the lost revenue can be reduced. Thus with a better understanding of these characteristics, a customized approach can be developed by the bank in the context of their Customer Relationship Management strategy.

VI. REFERENCES

[1] Sumathi S., Sivanandam S.N., "Introduction to Data Mining and Its Apllications", Springer Berlin Heidelberg, 2009.

[2] Md. Rashid Farooqi and Khalid Raza, "A Comprehensive Study of CRM through Data Mining Techniques" in proceedings of the National Conference: NCCIST-2011, September 2011.

[3] Kanti Das Sanjay, "Customer relationship management in banking sector", Arth Prabhand: A Journal of Economics and Management, Vol.1 Issue 6, pp. 68-82 September 2012.

[4] K. Tsiptsis and A. Chorianopoulos, "Data Mining Techniques in CRM: Inside Customer Segmentation", John Wiley & Sons, Ltd, pp. 1-13, 2009.

[5] Mutanen Teemu, "Customer churn analysis – a case study", pp.1-19 , march 2006.

[6] Bhambri Vivek, “Data Mining as a Tool to Predict Churn Behavior of Customers”in preceding of International Journal of Computer & Organization Trends, Volume 2 Issue3, ,2012.

[7] Dr. Devi Prasad .U , S Madhavi, “ Prediction Of Churn Behavior Of Bank Customers Using Data Mining Tools”, Business Intelligence Journal, Vol.5 , pp.97-101, January, 2012.

[8] Chitra .K, Subashini .B , “ Customer Retention in Banking Sector using Predictive Data Mining Technique”, in preceding of 5th International Conference on Information Technology ,2011.

[9] Margaret H. Dunham , “Data Mining : Introductory and Advanced Topics”, Pearson Education,2005

[10] Srinivasan G, “Services Sector and its Contribution to the Indian Economy”, Yojana, Volume 55, pp. 5-7, September 2011

[11] www.cs.waikato.ac.nz/ml/weka, “Machine Learning Group at the University of Waikato”.