Prediction Of Churn Behaviour Of Bank Customers Using Data Mining Tools

**** U. Devi Prasad***
***** S. Madhavi***

INTRODUCTION

Data mining is evolving into a strategically important dimension for many business organizations, including the banking sector. It is a method of analyzing the data from different viewpoints and summarizing it into valuable information. Data mining assists the banks to look for hidden pattern in a group and discover unknown relationship in the data (Fathimathabasum). Data mining can contribute to solving business problems in banking and finance by finding patterns, casualties, and correlations in business information and market prices that are not immediately apparent to managers, because the volume of data is too large or is generated too quickly to be screened by experts (Dass). Data mining techniques help companies, particularly banking, telecommunication, insurance, and retailing companies to build accurate customer profiles based on customer behavior. Thus, it is becoming a necessity in this competitive environment to analyze the data from data warehouse containing hundreds of gigabytes or terabytes of data (Fathimathabasum). Data mining tools predict patterns, future trends and behaviors, allowing businesses to effect proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analysis of past events provided by retrospective tools typical of decision support systems.

The importance of collecting and analyzing data to achieve a competitive advantage is widely recognized in today's age of information. Modeling and investigating a system and discovering relations that connect variables in a database is the objective of data mining (Berson et al., 1999). Data mining uses different models for the creation of information about data, which is known as the discovery model. Data mining uses methodologies that can sift through the data in search of frequently occurring patterns, can detect trends, produce generalizations about the data, etc. These tools can discover these types of information with very little (or no) guidance from the user (Weiss,1997). The main tasks are Prediction, Classification, Detection of relations, Explicit modeling, Clustering, and Deviation Detection. Moreover, since the data mining process is systematic, it offers firms the ability to discover hidden patterns in their data-patterns that can help them understand customer behavior and market trends.

AN ILLUSTRATIVE DATA MINING APPLICATION IN BANKING: CHURN MODELING

For forecasting the future to churn, a very vigorous model should be in hand, and an active model can only be built if we have a vigorous dataset in hand. Hence, data preparation is a vital step in churn prediction, and it takes almost 60-70 percent of the total time. In this part of the paper, the researchers make clear the kind of data required for an active system, commonly available customer data in the real life, guidelines to obtain useful attributes from the obtainable data. Constructing a model for churn prediction means that we are trying to model the customer's behavior (churning out). For this to be successful, the customer transaction activities should be analyzed in a specific period of time. Hence, taking the data would never be enough for the requirement. On the other hand, considering the transaction activities in a fixed time period would also not satisfy the requirement. The reason can be explained by an example. Say, for example, a model is built by using the data of 1000 customers, of which 700 are active, and 300 are known to be churned out, and their 3 months activities are analyzed (say, from February, 2006 to April, 2006). Here, the time period is fixed and the activities done in this time period of all the 1000 customers were only analyzed. Now, out of the 300 churn customers, say 50 percent of them churned away in February. This means, the model will not be fully trained with the behavior of churn customers before churning, as only one month's activity was analyzed. This problem

* *Associate Professor,* Hyderabad Business School, GITAM University, Hyderabad. E- mail : prasad_vungarala@yahoo.co.in
** *Assistant Professor,* Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh.
E-mail : sripathi.madhavi235@gmail.com

occurred because of fixing the timeline beforehand as shown in the Figure 1. In this paper, the researchers considered a dynamic time period, which differed for each customer. This concept would be better explained by continuing the above example. If a customer churned away in Feb, 2006, from that point of rime, the past 3 months activity is considered i.e., transaction activities done in Dec 2005, Jan 2006, and Feb 2006. And if another customer churned away in March 2006, transaction activity of Jan 2006, Feb 2006, and March 2006 should be considered. This can be seen in the Figure 2. In this way, the researchers adopted an active timeline for every customer, and hence refrained from the difficulty of not training the model accurately. This abstract idea applies to the information constituting an account of churn that has occurred. For diligent (active) records, the researchers can take into account (or examine) the behavior in any 3 months portion of time. In their detailed examination, the researchers considered the behavior of the most recent 3 months before the last transaction date of the active customers. The total count of months of data to be considered for churn analysis is a business problem. Usually, considering the transaction activities of 3 months would (be adequate) satisfy the requirement.

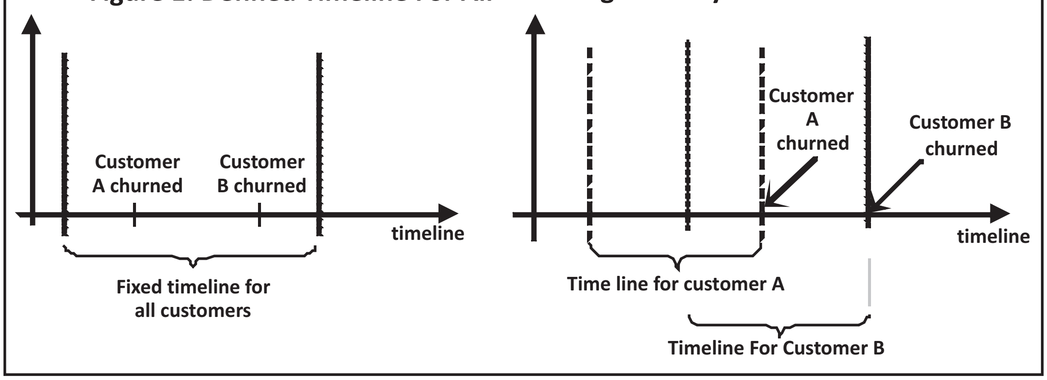

Figure 1: Defined Timeline For All **Figure 2: Dynamic Time Period**

Next, the researchers make clear about the actual data they used and elucidate about the data filtering steps to prepare an efficient dataset. The researchers acquired the customers' data from a Nationalized Indian Bank. The particulars of the data acquired are shown in the Table 1.

Table 1 : Details Of Customer Data

Table Name	Attributes	No. of Records
Customer	Custno, Name1, Name2, Address, Status, DoB (Date of Birth), Edn	40,870
General Ledger	Custno, AcNo, Descr, DOP (Date of a/c opening)	1,08,019
Dormant	Acno, Descr, Dormant	17,992
Master	Acno, Balance, Dormant flag	31,012
Txn	Acno, Trntype, Date, Amount	31,39,010
Ttype	Ttype#, Typecode, Descr	93

Source: Bank Records

The Table 1 shows the customers' details like customer number, name, address, date of birth, and status. Completely, there were 40,870 customers. The general ledger table holds the account numbers, account types, date of opening, and description of accounts. Here, there were 1, 08,019 accounts for the above defined informed as being dormant since a while. The master table holds all the account numbers and their latest balances. The Txn table holds the last 5 years' transactional details of all the accounts. Finally, Ttype table carries the description for different transaction types. In all, 93 transaction types had been outlined.

The final dataset was prepared on the basis of the available data shown in the Table 1, which contains the attributes such as: customer number, Duration (Dur), number of Credit transactions in 3 months (CRTxnx), number of Debit transactions in 3 months (DRTxns), Average credit amount in 3 months (AvgCrAmt), Average debit amount in 3 months (AvgDrAmt), total number of other accounts (NumOtherAccs), percentage of accounts closed in 3 months (PercClosedRecently), and status (status).

The Duration attribute consists of the number of months the customer transacted with the bank. When the researchers say that 3 months is an attribute, they refer to the concept of dynamic timeline explained previously. The final dataset consisted of 1,484 records. Out of which 1,163 were active customers, and 311 were churn customers. Extracting this dataset from the raw data available was not an easy job. Further in this section, the researchers share their experiences in working with the above data, and subsequently extracting the data for the required set of attributes.

There are many discrepancies in data. In these discrepancies, the most important and the most crucial one is the status attribute in the customer table, which is explained here. The status attribute describes whether a particular customer is active, inactive, or is churned out. So, the status attribute develops the class variable in the dataset. A customer may have more than one account with the bank. Then, the difficulty here was if one of the accounts of the customer remained dormant, then that condition is set as churned out. However, the customer is still with the bank as his/her other accounts are still active. Therefore, it was realized that, it was very difficult to prepare the training dataset based on the status attribute. However, the status of each & every account was very much required to develop the class variable of the dataset. The Descr attribute in the general ledger table served to verdict the status of each account. The Descr attribute consists of the description of the account type and moreover, that whether the account is inactive or is still active. Hence, by using the Descr attribute in the general ledger, the status attribute of the dataset was developed.

Another problem with the data was that some of the attributes had missing values. Fields like DOB, DOP were partially developed. Due to this reason, demographic attributes such as age and gender could not be considered in the final dataset.

The data consists of different types of accounts like savings account, current account, cash credit account, loan account etc. In the analysis, the behavior of savings account customers was analyzed. There were 22,155 savings accounts out of 1, 08,019 total accounts. Out of these 22, 155 accounts, 6, 633 accounts were found to be inactive from the dormant table. But here, the researchers were not concentrating on modeling the behavior of dormant customers, so these accounts were ignored from the target customer base. There were some accounts whose duration of transacting with the bank was not much i.e., number of months transacted with the bank were very less. Such accounts gave a notion that these accounts were opened for a particular purpose, and were closed as soon as that purpose was fulfilled. Considering these accounts may provide a poor dataset and consequently, lead to the creation of a despicable predicting model. So, the researchers ignored the records of those accounts whose duration of transaction was less than 6 months. There were some set of accounts whose duration of transaction was more than 6 months, but they had very less transaction activity. This gave a view that these customers had just opened the savings accounts and rarely carried out transactions through them. Such type of data also causes poor modeling. Hence, the records whose number of transactions were less than 50 were also ignored from the target customer base. After operating all these filtering steps, the target customer base was reduced to 1, 484 accounts. Out of the 1,484 accounts, 1,163 were active records, and 311 were churn (inactive) records.

By originating the right sort of data from the obtainable raw data, 1,484 customers' transactional behavior was analyzed. As expressed before, in order to model the behavior of both active and churn customers, the researchers had to practice (built) the model with their most recent behavior. The most recent behavior of accounts could be obtained from the Txn table, which has the transactional details of all the accounts. Txn table holds the Trntype field, which describes the kind of the transaction involved such as credit voucher, deposit, inward cheque clearing, etc., Trntype field can take the values of credit voucher, cash deposit, inward cheque clearing etc. There were 93 different transaction types, and each could be recognized as either a credit transaction, or a debit transaction. After separating the credit and debit transactions done by the customers, the researchers had to calculate the number of credit transactions (CRTxn) and number of debit transactions (DBTxn) for all the 1,484 customers. The average amount of money transacted by the customers in the defined timeline may also provide support in training the model in a better manner. For this reason, two more attributes, say, the average amount included in credit transactions (AvgCRTxn) and the average amount included in debit transactions (AvgDBTxn) were intended for the 1,484 customers. As the researchers mentioned earlier, they had taken into account only the transaction behavior of the savings accounts of the customers. As these customers had other accounts with the bank, the number of other accounts and percentage of closed accounts in the defined timeline may also help to train the model better. So two more attributes, say, number of other accounts (NumOtherAcc), percentage of other closed accounts in the defined timeline of 3 months (PercClosedRecently) were intended for all the 1,484 customers.

CONSTRUCTING DATA MINING MODELS AND TRAILING OUTCOMES

In the previous part, the researchers explained about the data filtering and preparing steps in brief. After having the dataset in hand, the immediate step was to prepare a predictive model by using this dataset. Usually, Data mining methods are used to prepare data models, and these models consequently help for future predictions. As predicting churn is exclusively a classification problem, supervised data mining techniques are used to take away this problem. Here, the researchers used two classification tree algorithms, say, CART, C5:0 for preparing the two classification trees.

Table 2 : Confusion Matrix And Prediction Success Rate For Training Data

True Class	Total # samples	Predicted Active	Predicted Churn	Success per cent
Active	926	797	129	85.79
Churn	261	13	248	95.01

Table 3 : Confusion Matrix And Prediction Success Rate For Test Dataset

True Class	Total # samples	Predicted Active	Predicted Churn	Success per cent
Active	241	208	33	86.30
Churn	57	5	52	91.22

Table 4 : Decision Rules Generated By CART

Rule #	Rule	Predicted Class	# Cases
1	AvgDrAmt <= 608 and AvgCrAmt <= 37.5	Churn	93
2	AvgDrAmt <= 608 and AvgCrAmt > 37.5 and AvgCrAmt <= 1655.5 and Duration > 18	Active	61
3	AvgCrAmt <= 1655.5 and AvgDrAmt > 608	Churn	102
4	AvgCrAmt > 1655.5 and Duration <= 23.5 and AvgDrAmt <= 1300.5	Active	20
5	AvgCrAmt > 1655.5 and Duration <= 23.5 and AvgDrAmt > 1300.5 and PercClosedRecently > 0.0416667	Churn	26
6	AvgCrAmt > 1655.5 and Duration > 23.5 and Duration <= 27.5	Active	615
7	Duration > 27.5 and Duration <= 68.5 and AvgDrAmt <= 3421.5 and AvgCrAmt > 1655.5 and AvgCrAmt <= 3674	Churn	16
8	Duration > 27.5 and Duration <= 68.5 and AvgDrAmt <= 3421.5 and AvgCrAmt > 3674	Active	18
9	AvgCrAmt > 1655.5 and Duration > 27.5 and Duration <= 68.5 and AvgDrAmt > 3421.5	Churn	38
10	AvgCrAmt > 1655.5 and Duration > 68.5	Active	43
11	AvgDrAmt > 1300.5 and PercClosedRecently<= 0.04 and AvgCrAmt > 1655.5 and AvgCrAmt <= 17894.5 and Duration <= 17.5	Churn	70
12	PercClosedRecently <= 0.04 and Duration > 17.5 and Duration <= 23.5 and AvgDrAmt > 1300.5 and AvgDrAmt <= 7449 and AvgCrAmt > 3527 and AvgCrAmt <= 17894.5	Active	32
13	PercClosedRecently <= 0.04 and Duration > 17.5 and Duration <= 23.5 and AvgDrAmt > 1300.5 and AvgDrAmt <= 7449 and AvgCrAmt > 1655.5 and AvgCrAmt <= 3527	Churn	10
14	AvgDrAmt <= 608 and AvgCrAmt > 37.5 and AvgCrAmt <= 1655.5 and Duration <= 18	Churn	4
15	PercClosedRecently <= 0.04 and AvgCrAmt > 1655.5 and AvgCrAmt <= 17894.5 and Duration > 17.5 and Duration <= 23.5 and AvgDrAmt > 7449	Churn	10
16	Duration <= 23.5 and PercClosedRecently <= 0.04 and AvgCrAmt > 17894.5 and AvgDrAmt > 1300.5 and AvgDrAmt <= 14238.5	Active	10
17	Duration <= 23.5 and PercClosedRecently <= 0.04 and AvgCrAmt > 17894.5 and AvgDrAmt > 14238.5	Churn	10

CLASSIFICATION TREE MODEL USING CART

The Classification Tree is built by adopting the Classification and Regression Tree (CART) model on the training dataset with the following specifications: Optimal tree cannot be discovered since CART does not use the stopping rule. Thus, firstly, the tree is over grown, and is then pruned back to ensure that significant patterns are not overlooked by stopping too soon. The advantage with CART is that it performs binary splitting to make the data more sparing and to detect more patterns before too few data are left for learning. The study used Gini Concentration Coefficient to abridge power curves of prediction. The explanatory variables are Customer Duration, CRTxn, DRTxn, AvgCrAmt, AvgDrAmt, PerClosedRecently and the target variable is Status. The CART gives rules for the target variable as a function of other fields in the dataset that were previously identified as explanatory variables. 80 percent of the dataset i.e., 1,187 samples consisting of 926 active customer records and 261 churned customer records were considered in training the dataset. The remaining 20 per cent of the dataset i.e., 296 samples consisting of 241 active customer records, and 57 churned customer records were considered for testing the dataset. The confusion matrix and the prediction success rate of the training dataset and the testing dataset are shown in the Table 2 and Table 3 respectively.

Table 5 : Confusion Matrix And Prediction Success Rate For Training Data

True Class	Total # samples	Predicted Active	Predicted Churn	Success per cent
Active	926	881	45	95.14
Churn	261	60	181	69.3

Table 6 : Confusion Matrix And Prediction Success Rate For Test Data

True Class	Total # samples	Predicted Active	Predicted Churn	Success per cent
Active	241	232	9	96.26
Churn	57	18	39	68.4

The retention rate of active customers is comparatively less because, although some of the customers' status is marked as active, they exhibited churn characteristics. It is this segment of customers upon which the bank has to focus and apply churn prevention strategies. There were 17 leaf nodes in the tree model generated using CART, and hence, 17 decision rules could be drawn from it. The Table 4 depicts the 17 decision rules. Among the 17 rules generated by CART, 12 rules had adequate number of cases, and these rules can be adopted by the bank manager for predicting future churn customers.

CLASSIFICATION TREE MODEL USING C5.0

Another classification algorithm that produces decision trees with variable branches per node is C 5.0. Status is taken as the Target Variable and Customer Duration, CRTxns, DRTxns, AvgCrAmt, AvgDrAmt, PercClosedRecently are used as explanatory variables. 80 per cent of the dataset i.e., 1,187 samples consisting of 926 active customer records and 261 churned customer records, were taken in the training dataset. The remaining 20 per cent of the dataset i.e., 296 samples consisting of 241 active customer records and 57 churned customer records were taken in the testing dataset. The confusion matrix and prediction success rate of the training dataset and the testing dataset are shown in the Table 5 and Table 6 respectively.

DISCUSSION OF RESULTS

In this analysis, the researchers experimented with 2 classification techniques namely CART, and C 5.0 on 1,484 sample of bank customers, out of which 1,163 were active customers, and 311 were churn customers. The researchers used CART and C5.0 to trace out significant customer characteristics to predict churn. While CART yielded 95.01 per cent classification rate on training data, and 91.22 per cent on test data, C5.0 yielded 69.3 per cent classification rate on training data, and 68.9 per cent on test data. The prediction success rate of Churn class by CART was quite high, but C 5.0 exposed poor results in predicting churn customers. However, the prediction success rate of Active class by C 5.0 is more than the other technique. In order to have significant benefits, the model should be able to predict the churn behavior better. Thus, a model with higher prediction success rate of Churn class (i.e., CART) has to be chosen for reaping higher benefits. In all the decision tree models, all the explanatory attributes were found to be influencing the target variable, i.e., the status of the customer.

CONCLUSION

In recent years, data mining has gained widespread attention and increasing popularity in the commercial world. Transforming raw customer data into information is the goal of data mining projects. However, failure to turn this useful information into customer satisfaction and increased profits is the key to why many such projects often fall short of expectations. Thus, it is essential to examine why the customers, as indicated by the churn model, are churned out. If the churn prevention program is effective, the bank can look forward to reaping significant benefits from its efforts. A company that can retain 5 per cent of its current customers can raise its profits by 25 per cent. In this paper, the researchers 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. They have explained the concept of dynamic timeline that should be considered while converting raw data into useful data. The researchers also extracted the data for the chosen attributes from raw customer data for a chosen set of 1,484 customers. Out of these 1,484 samples, 1,163 customers' status was active, and 311 customers had the status of churn. The researchers used CART and C5.0 to recognize significant customer characteristics to predict churn. While CART yielded 95.01 per cent classification rate on training data, and 91.22 per cent on test data, C5.0 yielded 69.3 per cent classification rate on training data, and 68.4 per cent on test data. The study predicts the future churn of banking customers that can be checked, 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; bank managers can develop a customized approach to customer retention activities within the context of their Customer Relationship Management efforts.

REFERENCES

1) Athanassopoulos, A.D. (2000). "Customer Satisfaction Cues To Support Market Segmentation And Explain Switching Behavior." *Journal of Business Research*, Volume 47, Issue 3, pp. 191- 207.

2) Berry M.J.A. & Linoff G. (1999). *"Mastering Data Mining: The Art and Science of Customer Relationship Management."* John Wiley and Sons, Inc. New York, pp. 335-341.

3) Berson, A.,et al. (19991). *"Building Data Mining Applications For CRM."* McGraw-Hill, Rockefeller Center, New York, pp. 825-844.

4) Dass Rajanish (2006). *"Data Mining In Banking and Finance: A Note For Bankers."* Indian Institute of Management, Ahmedabad, Prentice-Hall of India, pp. 3-9.

5) Fayyad U., Grinstein G.E. & Wierse A. (2002). *"Information Visualization In Data Mining and Knowledge Discovery."* Morgan Kaufmann Publishers. California, Academic Press.

6) Gehrke J., Ganti V., Ramakrishnan R., and Loh W. Y. (1999). *"BOAT - Optimistic Decision Tree Construction."* in Proc. ACM SIGMOD International Conference On Management of Data, Philadelphia, PA, pp. 169 - 180.

7) Han Jiewei, Kamber Michelin (2006). *"Data Mining: Concepts & Techniques."* 2nd Edition, The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers, Massachusetts, pp. 224-256.

8) Indiastudychannel.com, *"Fathimathabasum: Data Mining In Banks."*, http://www.indiastudychannel.com/projects/1804-Data-Mining-anks.aspx. accessed on February 22, 2011.

9) J. Gehrke, Ramakrishnan R., and Ganti V. (1998). *"Rain Forest A Framework For Fast Decision Tree Construction Of Large Datasets."* in Proc. 24th International Conference On Very Large Data Bases, New York, pp. 416 - 427.

10) Kotler, P. , Keller, K.L (2006). *"Marketing Management."* 12th Edition, Pearson Prentice Hall, Upper Saddle River, N.J., Northwestern University, pp.177-192.

11) Poel, Van Den, D. & Lariviere, B. (2003). "Customer Attrition Analysis For Financial Services Using Proportional Hazard Models." *European Journal of Operational Research*, Vol. 157, Issue 1, pp. 196-217.

12) Pressler Margaret (2003). *"Signs of Fraud Go Beyond Signature."* Washington Post, Washington, p. H05.

13) Quinlan J. R. (1987). *"Decision Trees As Probabilistic Classifiers."* in Proc., 4th Int. Workshop Machine Learning, Irvine, CA, pp. 31 - 37.

14) Rastogi R. and Shim K. (1998). *"Public: A decision Tree Classifier That Integrates Building And Pruning."* in Proc. 24th Int. Conf. Very Large Data Bases, NewYork, 1998, pp. 404 - 415.

15) Reichheld, F.F. and Sasser, W.E. (1990). "Zero Defections: Quality Comes To Service." *Harvard Business Review*, Vol. 69, September-October, pp. 105- 111.

16) Shafer J., Agrawal R., and Mehta M. (1996). *"SPRINT: A Scalable Parallel Classifier For Data Mining."* in Proc. 22nd International Conference On Very Large Data Bases, Mumbai , pp. 544 - 555.

17) Smith A. (2006). "CRM And Customer Service: Strategic Asset Or Corporate Overhead?" *Handbook of Business Strategy*, Vol .7, pp. 87 - 93.

18) Turban, E., Mclean, E., Wetherbe, J., Bolloju, N., Davison, R. (2002). *"Information Technology For Management: Transforming Business In The Digital Economy."* John Wiley and Sons, INC: New York, pp. 441-458.

19) Weiss, S.M. (1997). *"Predictive Data Mining: A Practical Guide."* Morgan Kaufmann Publishers, Massachusetts.