Churn Prediction in Telecommunication Using Data Mining Technology

Rahul J. Jadhav[1]

Bharati Vidyapeeth Deemed University,Pune
Yashwantrao Mohite Institute of Management, Karad,
Maharashtra INDIA
rjjmail@rediffmail.com

Usharani T. Pawar[2]

Shivaji University Kolhapur
S.G.M college Department of Computer Science Karad,
Maharashtra INDIA
usharanipawar@rediffmail.com

Abstract—**Since its inception, the field of Data Mining and Knowledge Discovery from Databases has been driven by the need to solve practical problems. In this paper an attempt is made to build a decision support system using data mining technology for churn prediction in Telecommunication Company. Telecommunication companies face considerable loss of revenue, because some of the customers who are at risk of leaving a company. Increasing such customers, becoming crucial problem for any telecommunication company. As the size of the organization increases such cases also increases, which makes it difficult to manage, such alarming conditions by a routine information system. Hence, needed is highly sophisticated customized and advanced decision support system. In this paper, process of designing such a decision support system through data mining technique is described. The proposed model is capable of predicting customers churn behavior well in advance.**

Keywords- churn prediction, data mining, Decision support system, churn behavior.

I. INTRODUCTION

The biggest revenue leakages in the telecom industry are increasing customers churn behavior. Such customers create an undesired and unnecessary financial burden on the company. This financial burden results in to huge loss of the company and ultimately may lead to sickness of the company, detecting such customers well in advance is an objective of this research paper.

II. DATA MINING IN TELECOMMUNICATION

In telecommunication sector data mining is applied for various purposes. Data mining can be used in following ways:

A. Churn prediction:

Prediction of customers who are at risk of leaving a company is called as churn prediction in telecommunication. The company should focus on such customers and make every effort to retain them. This application is very important because it is less expensive to retain a customer than acquire a new.

B. Insolvency prediction:

Increasing due bills are becoming crucial problem for any telecommunication company. Because of the high competition in the telecommunication market, companies cannot afford the cost of insolvency. To detect such insolvent customer's data mining technique can be applied. Customers who will refuse to pay their bills can be predicted well in advance with the help of data mining technique.

C. Fraud Detection:

Fraud is very costly activity for the telecommunication industry; therefore companies should try to identify fraudulent users and their usage patterns.

III. CHURN PREDICTION IN TELECOMMUNICATION

Major concern in customer relationship management in telecommunications 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.

The objectives of the application to be presented here were to find out which types of customers of a telecommunications company is likely to churn, and when.

In many areas statistical methods has been applied for churn prediction. But in the last few years the use of data mining techniques for the churn prediction has become very popular in telecom industry. Statistical approaches are often limited in scope and capacity. In response to this need, data mining techniques are being used providing proven decision support system based on advanced techniques.

The BSNL. Satara like other telecom companies suffer from churning customers who use the provided services without paying their dues. It provides many services like the Internet, fax, post and pre-paid mobile phones and fixed phones, the researchers would like to focus only on postpaid phones with respect to churn prediction, which is the purpose of this research work.

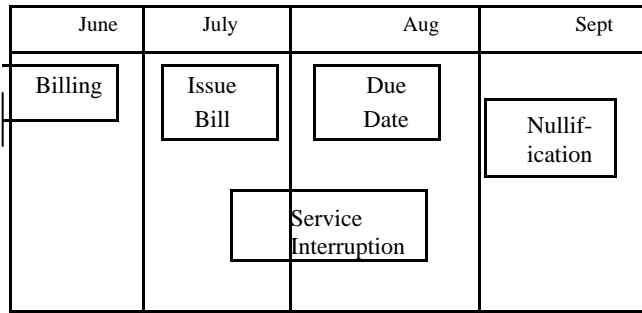| June	July	Aug	Sept
Billing	Issue Bill	Due Date	Nullif-ication
		Service Interruption	

Figure 1

As described in above figure1, customers use their phone for a period of one month, called the billing period. The bill is issued two weeks after the billing period. The due date for the payment is normally two weeks after the date of issue. If a bill is not paid in this period, the company takes action on such a customer's. The company disconnects the phone one way, two week after payment due date for 30 days. That means the customer can only receive incoming calls and can't make outgoing calls during these 30 days.

If the customer pays their bill, connection is reestablished. If the customer doesn't pay in this 30 days period the companies nullify the contract and uncollected amount will be passed to custody. The amount that customer owes is transferred to uncollectible debts and the company considers the money most probably lost. Telecommunication companies face considerable loss of revenue, because some of the customers who are at risk of leaving a company. As one can see the measures that the company takes against churn customers come quite late Predicting such customers well in advance who are at risk of leaving a company. Detection of as many such customers well in advance is the main objective of this research paper.

A. Data collection

Following are the different sources used for collecting the data.

In house customer databases- It has major fields such as phone number, address category, type of security deposit and cancellation.

External sources- Call detail record of every call made by the customer i.e. call no, receiver no., call date, call time and duration of each call. In addition data was identified from billing sections.

Research survey- Data is collected through previous research survey.

In order to make study more precise customers from various categories such as government, businesses and private were included. The following table shows the number of records from various categories were included.

Category	Number of records
Government	35
Business	125
Private	735
Total	895

Table 1: category wise records

B. Data Preparation

Before data can be used for data mining they need to be cleaned and prepared in required format. Initially multiple sources of data is combined under common key. Typical missing values on the call detail records like call_date, call_time, and call_duration was found, it forced to ignore such records in the study. At this stage, two attributes such as late_pay, and extra_charges were eliminated since records in these attributes were not complete even though they were playing significant role in the problem of churn. In order to perform above tasks SQL server were used.

Attribute	Data Type	Description
category_Id	Text	Category id of phone account.
Category	Text	Category of phone account.
Sec_dep	Currency	Security deposit

Table 2: Attributes from customer file

Attribute	Data Type	Description
Late_pay	Number	Count of late pay
Extra_charges	Number	Count of bills with extra charges.

Table 3: Attributes from billing information section

Attribute	Data Type	Description
Max_Units	Currency	Maximum no of units charged in any two week period during the study period.
Min_Units	Currency	Minimum no of units charged in any two week period during the study period.
Max_dur	Number	Maximum total duration for the calls in any two week period during the study period.
Min_dur	Number	Minimum total duration for the calls in any two week period during the study period.
Max_count	Number	Maximum no. of calls in any two week period during the study period.
Min_count	Number	Minimum no. of calls in any two week period during the study period.
Max_dif	Number	Maximum no. of different numbers are called in any two week period during the study period.
Min_dif	Number	Minimum no. of different numbers are called in any two week period during the study period

Table 4: Attributes from call detail record

C. Defining data mining function

Churn prediction can be viewed as a classification problem, where each customer is classified in one of the two classes such as most possible churning or not. Even though there were many

churning customers reported, it was difficult to get a significant number of them during study period. As a result, the distribution of customers between the two classes was very uneven in the original dataset. Approximately 82.33% were not churning and 17.67% churning customers during research period. Classification problem with such characteristics are difficult to solve. Hence new dataset had to be created especially for the data mining function. For every phone call made by a customer, the data had to be aggregated. Aggregation is done with the aim of creating a customer profile that reflects the customer's phoning behavior over the last five months. The details of this aggregation process are complex, interesting and important, but cannot be described here due to space limitations. In essence, many aggregated attributes containing the lengths of calls made by every customer in these five months were created.
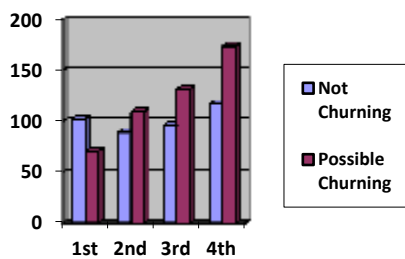
	1st week	2nd week	3rd week	4th week
Not Churning	102.5	89.2	96.1	117.2
Possible Churning	71	109.5	131.5	174

Figure 2: Difference between Churning and not churning customers

From the above figure the difference between the not churning and possible churning customers can see clearly. On an average, the not churning customers were using their phone approximately for the same number of times during all periods ranging from 102 to 117. On the contrary, the possible churning customers on the average were using their phone for less number of times for the first few days and then their behavior changed resulting to high number of calls than not churning customers, ranging from 71 to 174.

D. Model Building and Evaluation

The major task to be performed at this stage was creating and training a decision support system that can discriminate between churning and not churning customers. For the proposed, we had a choice of several data mining tools available, and METALAB was found to be the most suitable for this purpose, because it supports with many algorithms. It has neural network toolbox. The algorithm that is used for this research work is Back propagation algorithm. While building a model whole dataset was divided into three subsets. These subsets were the training set, the validation set, and the test set. The training set is used to train the network. The validation set is used to monitor the error during the training process. The test set is used to compare the performance of the model.

IV. CONCLUSION

This research report is about to predicting customers who are at risk of leaving a company, in telecommunication sector. Using this report company will be able to find such kind of customers. The model can be employed in its present state. With further work, the scope of model can be widened to include insolvency prediction of telecommunication customers.

REFERENCES

[i] Abbot D. " An evaluation of high end data mining for fraud detection" 1996 www.kdnuggets.com

[ii] Berthold and Hand / Hardcover / 2003 (Second Edition) "Intelligent Data Analysid"

[iii] Bigus J. P. "Data mining with neural network" McGraw-Hill Dalkalaki S. " Data mining for decision support on customer insolvency in telecommunication businesses" 2002 URL: www.elsevier.com/locate/dsw

[iv] Issac F. "Risk and Fraud management for Telecommunication" 2003.

[v] Kurth thearling Tutorial "An Introduction to Data Mining www.thearling.com

[vi] Zhaohui Tang and Jemie Maclennan "Data mining with SQL Server 2005, wiley.

[vii] Michael Berry & Gordon Linoff / Paperback / 1999 "Mastering Data Mining"

[viii] Data mining techniques(2nd edition), Michael J. A. Berry and Gordon S. Linoff, Wiley India

[ix] Data ware housing Data mining and OLAP ,Alex berson and Stephen Smith

[x] Data mining A tutorial based premier, Richard J. Roiger, Michael Geatz.

[xi] Business Intelligence and Insurance, White Paper, Wipro Technologies, Bangalore,2001

[xii] Revenue Recovering With Insolvency Prevention On a Brazilian Telecom Operator" (Calos Andre R. Pinheiro, Alexander **G.** Evsukoof, Nelson F. F. Ebecken Brazil)