Introduction

“Churn” is a common phenomenon that occurs in telecom Industry. By “Churn” we mean those customers, who will be leaving us in near future. If we are able to predict in advance, the attributes of customers whom we are going to loose in near future one can take corrective action so that we can minimize this phenomenon.

As we are aware that acquiring a new customer is more expensive than retaining an existing customer. In fact it costs anywhere between 6 to 10 times more for acquiring a new customer. Also one bad customer can in fact spoil the chances of acquiring 8 – 10 good customers.

Initially when the industry growth is high, churn may not be a problem. The focus at this stage is on customer acquisition. For each customer churning out there may be many who will be joining. As the industry matures, the churn rate rises and there will probably be a time when for every one customer leaving only one new will be joining or there may not be even one joining.

When this stage is reached, it will affect the bottom line for the company.

So predicting the churn is a very important step for keeping the bottom line of the company in a competitive environment.

Also in a mature market growth comes from:
✓ Maximizing profit from existing customers (cross selling and up selling)
✓ Retention of customers
✓ Stealing new customers from its competitors.

Some unique features of Telecom Industry

In many respects telecom Industry is like any other service Industry like Financial, Insurance & utilities. But it has some peculiar features like high cost of acquisition, not much of face – face contact, little customer mind care. Compared to some other industries there is tremendous amount of data available as each transaction of the customer is recorded and data is in digital Format.

The problem:

One of a well known mobile phone company operating in the country for the past two years is interested to make investment in CRM technology. The market is becoming more and more competitive because of deregulation and slowly the industry will reach a maturity stage.

The maturing of the market and the increasing competition, the company wants to focus on its existing customers, how to keep them and make them more profitable.

What benefit such a model helps
One of the important benefits of such as model will be to marketing department /call center(s). The list of customers who are likely to churn in the coming month will be made available to the above places so that further action can be taken. Data analyzed from back end or Analytical CRM System will go to front end or operational part of CRM closing the CRM loop. Marketing department
can plan giving customers discounts, other promotions/events, other products of other sister companies wherever applicable.

Predicting the churn also helps us to approximately know the life time value of customers. If a group of customers have a 20% chance of churning this month, then we would expect them to remain customers for 5 months. (1 month ÷ 20%). If the churn were reduced to 1%, then we would expect the customers to remain for 100 months.

The other application is for prioritizing customer segments. If a segment is more likely to churn, perhaps, they should not get a high value gift. May be a discount might encourage them to stay. The issue may not be clear cut, but having a churn score will definitely help in better Decision making.

HOW DO WE GO ABOUT CHURN PREDICTION?

Data requirements for the Analysis

The basic requirements are:

✓ Data from customer information file like age, sex, Zip code etc.
✓ Data from service account file such as Pricing plan, activation data, contract identification etc.
✓ Data from billing system such as number of calls, airtime, fixed line time, total amount spent, no. of times calls made to customer care center, change in price plan etc.

ANALYZING THE DATA

To analyze the churn there are different techniques available. Some of the common models used are:

✓ Regression analysis
✓ Decision Trees
Out of the above models, Decision Trees are a good choice since they provide the Rules that Business users can understand. Other models like Neural Networks, Logistic Regression reduce the understandability of the phenomenon even though they may be little bit more accurate than decision trees.

Sometimes important derived variables like growth rate of the number of calls over a period of time, calls proportions changes etc. are included. One rule of thumb of including derived variables are those that will explain the phenomenon of real world rather than including mere mathematical transformations.

Building a churn model:
Let us build a Churn Model for a Telecom company. Using the data provided in the dataset Telecom.xls. Though real life example may have many more variables and volume of data will be extremely large nevertheless this case indicates the methodology of model development.

The variables in the data set are:
- Contract id
- Phone number
- Age (since how long he is the customer)
- Change in price plan
- Air charge
- MTNL charge
- Total call duration
- Total charges
- No. of incoming calls
- No. of outgoing calls
The method of construction of data set form operational system is as follows:

Assume that we are building a model to predict those customers who are likely to leave in the month of September 2002. We take the sample data set and this set consists of those customers who are continuing in the month of September 2002 and those who have left in the month of September 2002. Those who have left have been coded as “Churned” and those who have not left have been coded as “continuing”, irrespective of which day of the month they have left. For each of these customers, we go back three months in time period i.e. for August 2002, July 2002, June 2002. Three months of data will be used to build the model. Each month will be divided into two windows of 15 days size. So we will have six windows which are called w1,w2,w3,w4,w5 and w6. w1 is window for June 1 – 15, w2 for June 16 – 30 etc.

Each of the above variables from operational database is extracted, summarized for each contract number. In that sense, each contract number can be treated as equivalent to a customer. So after summarization, there will be one record per customer containing data for all the above fields you can see this by opening the excel file “Telecom.xls”.

The concept is illustrated pictorially as follows:
For this illustration we will use status as target variable and all others as input variables. (If you are sure that some variables are not useful, you can leave them from the analysis).

Use FORESIGHT to build the tree. (up to the depth of 3).

The details of building the tree are given in the software manual.

Since contract _Number and phone-number are unique, they should not be included to develop the model. First build a normal tree with default values. Use 20% of the data to test the model.

INTERPRETATION OF RESULTS:

The training data set contains 3715 samples (80%). Balance is Test set (20%) which will be used to validate the model. The entire training set contains 37% churners.

The first most important variable which has highest impact on churners is change in price plan (ch_plan). The tree has made two Groups; one group for which change in plan is \( \leq 2 \) (group no. 1) and for another \( > 2 \) (Group no. 2). The first Group consists of 30% churners and 70% non-churners. While the second group consists of 94% churners and 6% non churners. So it is important we concentrate on customers in Group 2 and design an intervention program to retain them.

Now let us concentrate on Group no. 2. By double clicking this group, this node is split into Two group. One is the group where w6TOTAL CHARGES \( \leq 1187 \) and another w6TOTAL CHARGES \( > 1187 \). 96% of the churners are in group where w6TOTAL CHARGES are \( < 1187 \).
So the next important attribute of churners is total charges in the window preceding the month of churn.

Now let us come to the segment where change in plan is less than 2. Drill down this segment by double clicking. Now the segment is broken into two groups: one Group where W5outcalls are less than \( \leq 0 \). 87% of the churners are in group where W5outcalls are <0.

So we can create business rule like:

1. If change in plan \( \leq 2 \) and
   \[ \text{W5outgoing calls} \leq 0 \] then 87%
2. If change in plan >2 and
   \[ \text{W6totalcharges} < \text{Rs 1187} \]
   Then the group has 96% churners

We can drill further down for details. Normally in many practical applications up to 5 levels should be sufficient to draw meaningful conclusions.

SECOND ANALYSIS:

Let us build a model to analyze the churn in top 25% of segment.

We define top 25% of customers based on total bill amount for the month of June, July and August 2002. to extract this data do back to the earlier excel sheet and use autofilter option of excel to select top 25%. Save this file as telecom data top 25 percent. Use this data set for further analysis.

Using FORESIGHT open the file for analysis. Since we have segmented the data set using total bill amount this can be excluded from the analysis, the other field(s) that can be excluded from the analysis are Last2TOT, TOTAIR and RATL2AIRC.
As usual target variable is STATUS. Select 20% test data. This time change the minimum number of outcome to 50 and percent of pruning to 10. The above parameters will have effect on the depth of tree. This ensures that sum of samples in any parts of the tree is 50.

Results Interpretations:
For understanding of the results, let us analyze the tree to a depth of Two/Three levels.

The first important variable which will help to predict the churn is change in price plan. Whenever change in price plan is greater than TWO almost all the customers in this segment have churned out. (Churn rate 100%).

The second important group is those whose change in plan is less than or equal to Two and in Window6 (i.e. Fifteen days window prior to prediction period). MTNL charges less than Rs. 7.20. This segment also has a churn rate of 100%.

As an exercise compare the churn Groups characterizes of all customers Vs those in Top 25% based to total amount.

CONCLUSIONS
The above case study illustrates how an insight into the process of “Customer Churn” can be gained using the Analytical CRM techniques. It also tells the management what are the important variables that affect the churn. This will help to put a churn management system in place so that early information about those customers can be obtained.

The output of churn Management system can be sued by marketing department to develop suitable strategies/ interventions.
HANDS ON EXERCISES

1. Select bottom 25% of customers from the Telecom.xls files and built a decision Tree to get an Insight into attributes of “churners”.

2. How this “churners group” compare with those of “churners Group” in Top 25%

3. What is the average age and average air charges of churners of Top 25% and bottom 25% of customers a the 2nd level of decision tree

4. The total number of calls made to the customer care center is given in second file Customercare.xls

Merge this information to the original data file segment the data into two groups.

   I. Those who have not made any calls to the customer care center.

   II. Those who have made calls

Build churn prediction model to each group separately

Discuss the results

5. What additional / derived variables would you think can be collected to improve the accuracy of the model?