Telecom Churn Prediction with Main Drivers

Name:

Mr. Kapil Madan

Agenda

1. Background and Objective

2. Problem Solving Methodology

3. Definition of High Value customers and Churn

4. Overview of given customer base from data

5. Class Imbalance & SMOTE

6. Model building for Churn Prediction

7. Main Drivers for predicting customer Churn

8. Recommendation to manage customer churn

9. Appendix

Background and Objective

Background :

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, **customer retention** has now become even more important than customer acquisition.

To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**
In the Indian and the southeast Asian market, approximately 80% of revenue comes from the top 20% customers (called high-value customers). Thus, if they can reduce churn of the high-value customers, they will be able to reduce significant revenue leakage.

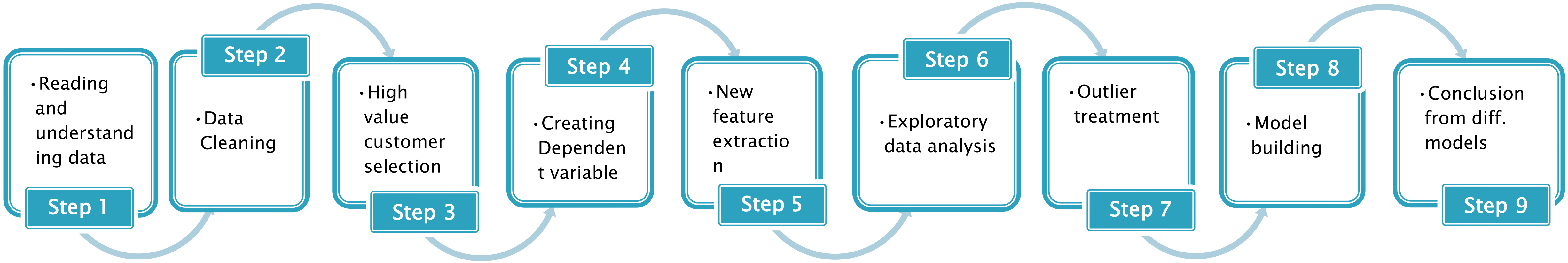
Objective:

- It is required to build a model to predict the High Value Customers (HVC) who can churn in recent future

- It will be used by the management to understand main drivers of churn. They can accordingly make their new strategy to retain the customers

- The model will be a good way for management to understand the requirement of the customer base.

Problem Solving Methodology



- Understanding data , Data cleaning and data preparation for model are most important parts of the analysis
- We have spend almost 60% time for those steps of the whole project
- Model building and evaluation of model took 40% time

Definition of High Value customers and Churn

High Value Customer (HVC):

In the Indian and the southeast Asian market, approximately 80% of revenue comes from the top 20% customers. Those customers are call High Value Customer.

For this project we have considered as below:
Those who have recharged with an amount more than or equal to X, where X is the 70th percentile of the average recharge amount in the first two months (the good phase).

1. Calculate the total recharge amount (call + data) for June & July
2. Average Recharge amount
3. Find 70^{th} percentile recharge value
4. Customers whose Avg. Recharge amt. > = 70 th percentile, considered as HVC

Customer Churn:

Customers who have not made any calls (either incoming or outgoing) and have not used mobile internet (2G or 3G) even once in the churn phase i.e. in September would be considered as churn and would be tagged as 1 otherwise customer would be tagged as 0

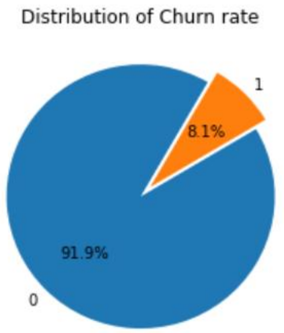
1. Calculate total incoming + outgoing minutes in churn phase
2. Calculate total 2G + 3G minutes in churn phase
3. If any customer did not use any calling or data services during Churn phased would be treated as churn

Overview of given customer base from data : **8.1% Churn** rate for HVC

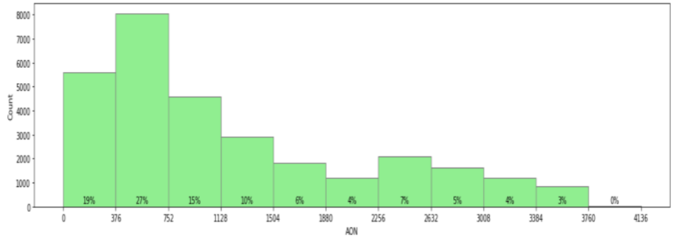
- During Good Phase, around 27% customer uses data(2G or 3G). Remaining 73% to 74% customers uses only voice call.

- Among 99,999 customers, around 30K customers are High Value Customers (HVC) who are spending at list 70 percentile of average spend for voice call and data

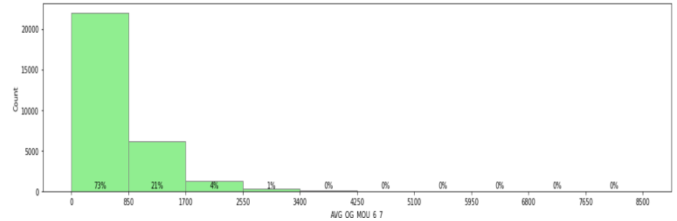
- Churn rate for HVC customers is 8.1% . Clearly **Class Imbalance** is there in HVC data.



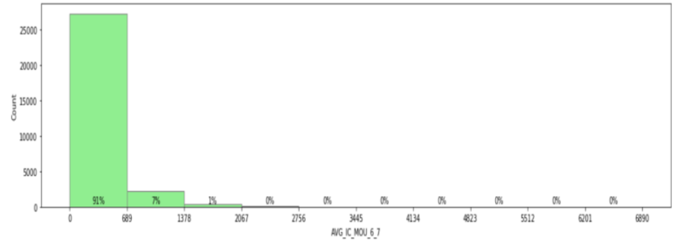
- Customer attachment with same network (AON):
  - 19% HVC customer are using the current network only for last one year
  - 27% HVC customer are using around 2 years
  - 29% HVC customers are attached with the network provider for more than 4 years(>1504 days)



- Total Outgoing calls:
  - 73% HVC customer's total outgoing calls during good phase (avg. of June and July)is around 850 minutes.
  - 21% HVC customer's total outgoing uses 1700 minutes



- Total Incoming calls:
  - 91% HVC customer's total incoming calls during good phase (avg. of June and July)is around 690 minutes.
  - 7% HVC customer's total incoming uses 1380 minutes



Class Imbalance & SMOTE

What is Class Imbalance?

when the number of observations belonging to one class is significantly lower than those belonging to the other classes in the data, then that problem is known as class imbalance/imbalance data.

Problem due to Class Imbalance?

Standard classifier algorithms like Decision Tree and Logistic Regression have a bias towards classes which have number of instances. They tend to only predict the majority class data. The features of the minority class are treated as noise and are often ignored. Thus, there is a high probability of misclassification of the minority class as compared to the majority class. So it is better to handle class imbalance before using those algorithms.

In our data 8.1% is churn customer. So we need to handle the class imbalance of the data. We have handle that by Synthetic Minority Over-sampling Technique (SMOTE) technique.

Synthetic Minority Over-sampling Technique (SMOTE) technique:

This technique works in below way:
A subset of data is taken from the minority class as an example and then new synthetic similar instances are created. These synthetic instances are then added to the original dataset. The new dataset is used as a sample to train the classification models.

Note: We should not use SMOTE on test data. Otherwisw test data would be inflated with synthetic records. We should validate the model only on original data¶

Model building for Churn Prediction:

Logistic Regression with ~ 85% recall value is final model

Dimensionality of data was quite high and many of the variables were correlated, so we have used Principal Component Analysis to reduce the dimensionality. Later we have build couple of classification model to predict churn accurately

i) **Logistic Regression:**
Training Accuracy: 84.04% , Training Recall: 85.08%
Test Accuracy: 83.77%, Test recall: 82.26%

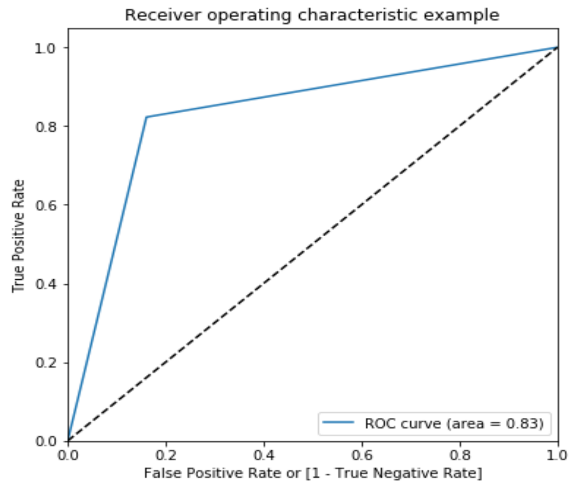
ii) **Decision Tree**
Training Accuracy: 86.56%, Training Recall: 87.33%
Test Accuracy: 82.91%, Test recall: 71.46%

ii) **Random Forest**
Training Accuracy: 85.66% , Training Recall: 84.36%
Test Accuracy: 85.86% Test recall: 74.66%

Selection of Final Model:

As our target is to find out the maximum number of high risk HVC accurately, we would give more importance on recall value. from that perspective we can see that Logistic Regression is performing better than other algorithms as both accuracy and recall value in train and test set are very near.

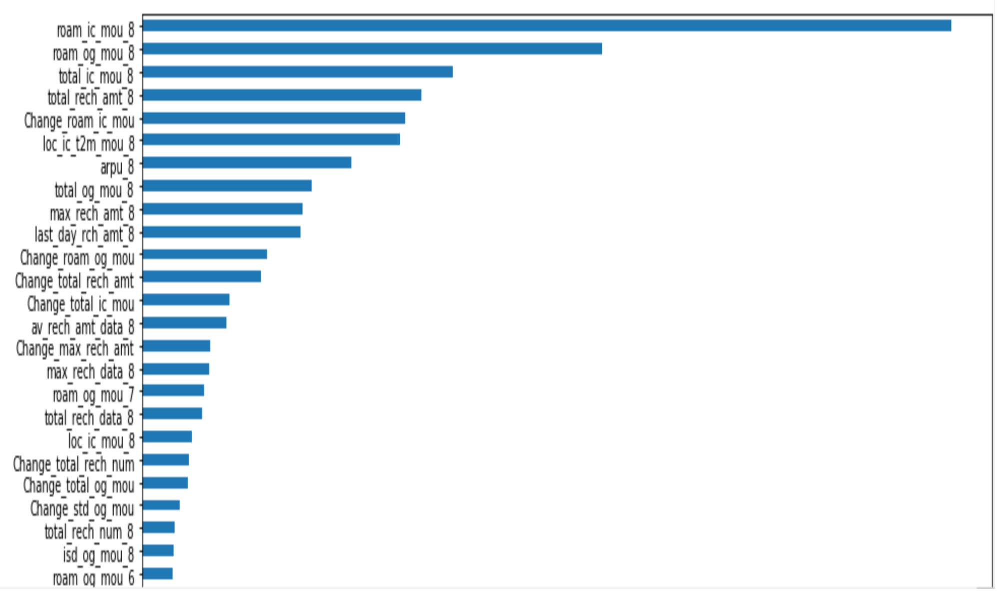
So we are selecting Logistic regression as our final model.



Main Drivers for predicting customer Churn

To understand the main driver of customer churn, we have build Random Forest model without using Principal Component Analysis.
Top 15 variables as per the importance of those variables are:

Features	Features Meaning
roam_ic_mou_8	Minutes of usage of roaming incoming voice calls in Aug
roam_og_mou_8	Minutes of usage of roaming outgoing voice calls in Aug
total_ic_mou_8	Minutes of uses of total incoming call in Aug
total_rech_amt_8	Total amt of recharge in Aug
Change_roam_ic_mou	Change of incoming roaming from good phase to Action phase (minutes)
loc_ic_t2m_mou_8	Minutes of usage of local incoming voice calls in other mobile network in Aug
arpu_8	Average revenue per user in Aug
total_og_mou_8	Minutes of usage of total outgoing voice calls in Aug
max_rech_amt_8	Max recharge amt in Aug
last_day_rch_amt_8	Last day recharge amount in Aug
Change_roam_og_mou	Change of outgoing roaming from good phase to Action phase (minutes)
Change_total_rech_amt	Change of total recharge amount from good phase to Action phase
Change_total_ic_mou	Change of total incoming usage from good phase to Action phase (minutes)
av_rech_amt_data_8	Average Recharge Amt in Aug
Change_max_rech_amt	Change of max recharge amount from good phase to Action phase



Recommendation to manage customer churn

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. Acquire a new customer is 5-10 times more costlier than to retain an existing one. Our recommendation to retain existing HVC are as follows:

- Telecom provider should monitor the features extracted by our Random forest model in previous page very carefully. From the EDA of total incoming and outgoing usage, we have noticed that behavior of non-churn customers are almost remain same during good phase as well as action phase. But for churn customers total incoming and outgoing usage drop drastically from good phase to action phase.

- Business should keep an eye of the changes of total voice call recharge amount, average recharge amount for data etc. During EDA, we have seen that those variables are dropping drastically for the churn customers during action phase.

- Telecom providers have to improve their services- specially for outgoing calls, data services etc.

Appendix

Top 20 variables as per feature importance of Random Forest Model:

	Features	Features Meaning
1	roam_ic_mou_8	Minutes of usage of roaming incoming voice calls in Aug
2	roam_og_mou_8	Minutes of usage of roaming outgoing voice calls in Aug
3	total_ic_mou_8	Minutes of uses of total incoming call in Aug
4	total_rech_amt_8	Total amt of recharge in Aug
5	Change_roam_ic_mou	Change of incoming roaming from good phase to Action phase (minutes)
6	loc_ic_t2m_mou_8	Minutes of usage of local incoming voice calls in other mobile network in Aug
7	arpu_8	Average revenue per user in Aug
8	total_og_mou_8	Minutes of usage of total outgoing voice calls in Aug
9	max_rech_amt_8	Max recharge amt in Aug
10	last_day_rch_amt_8	Last day recharge amount in Aug
11	Change_roam_og_mou	Change of outgoing roaming from good phase to Action phase (minutes)
12	Change_total_rech_amt	Change of total recharge amount from good phase to Action phase
13	Change_total_ic_mou	Change of total incoming usage from good phase to Action phase (minutes)
14	av_rech_amt_data_8	Average Recharge Amt in Aug
15	Change_max_rech_amt	Change of max recharge amount from good phase to Action phase
16	max_rech_data_8	Maximum Mobile internet recharge in Aug
17	roam_og_mou_7	Minutes of usage of roaming outgoing voice calls in July
18	total_rech_data_8	Total Mobile internet recharge in Aug
19	loc_ic_mou_8	Minutes of uses of local incoming call in Aug
20	Change_total_rech_num	Change of total recharge usage from good phase to Action phase

Thank You