Telecom Churn Case Study

By
Biji Krishna
Chitwan Tayal
Tanya Choudhary

OBJECTIVE:

The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well, understanding the typical customer behaviour during churn will be helpful.

Definitions of Churn:

There are various ways to define churn, such as:
Revenue-based churn: Customers who have not utilised any revenue-generating facilities such as mobile internet, outgoing calls, SMS etc. over a given period of time. One could also use aggregate metrics such as ‘customers who have generated less than INR 4 per month in total/average/median revenue’.
The main shortcoming of this definition is that there are customers who only receive calls/SMSes from their wage-earning counterparts, i.e. they don’t generate revenue but use the services. For example, many users in rural areas only receive calls from their wage-earning siblings in urban areas.

Usage-based churn: Customers who have not done any usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time.
A potential shortcoming of this definition is that when the customer has stopped using the services for a while, it may be too late to take any corrective actions to retain them. For e.g., if we define churn based on a ‘two-months zero usage’ period, predicting churn could be useless since by that time the customer would have already switched to another operator.

Understanding Customer Behaviour During Churn:

Customers usually do not decide to switch to another competitor instantly, but rather over a period of time (this is especially applicable to high-value customers). In churn prediction, we assume that there are three phases of customer lifecycle :

1.The 'good' phase: In this phase, the customer is happy with the service and behaves as usual.

2.The 'action' phase: The customer experience starts to sore in this phase, for e.g. he/she gets a compelling offer from a competitor, faces unjust charges, becomes unhappy with service quality etc. In this phase, the customer usually shows different behaviour than the 'good' months. Also, it is crucial to identify high-churn-risk customers in this phase, since some corrective actions can be taken at this point (such as matching the competitor's offer/improving the service quality etc.)

3.The 'churn' phase: In this phase, the customer is said to have churned. We define churn based on this phase. Also, it is important to note that at the time of prediction (i.e. the action months), this data is not available to us for prediction. Thus, after tagging churn as 1/0 based on this phase, we discard all data corresponding to this phase.
In this case, since we are working over a four-month window, the first two months are the 'good' phase, the third month is the 'action' phase, while the fourth month is the 'churn' phase.

Modelling:

Build models to predict churn. The predictive model that we are going to build will serve two purposes:

1.It will be used to predict whether a high-value customer will churn or not, in near future (i.e. churn phase). By knowing this, the company can take action steps such as providing special plans, discounts on recharge etc.
2.It will be used to identify important variables that are strong predictors of churn. These variables may also indicate why customers choose to switch to other networks.
In some cases, both of the above-stated goals can be achieved by a single machine learning model. But here, we have a large number of attributes, and thus we should try using a dimensionality reduction technique such as PCA and then build a predictive model. After PCA, we can use any classification model.

Also, since the rate of churn is typically low (about 5-10%, this is called class-imbalance) - we will try using techniques to handle class imbalance.

We can take the following suggestive steps to build the model:

1.Preprocess data (convert columns to appropriate formats, handle missing values, etc.)

2.Conduct appropriate exploratory analysis to extract useful insights (whether directly useful for business or for eventual modelling/feature engineering).

3.Derive new features.

4.Reduce the number of variables using PCA.

5.Train a variety of models, tune model hyperparameters, etc. (handle class imbalance using appropriate techniques).

6.Evaluate the models using appropriate evaluation metrics. Note that it is more important to identify churners than the non-churners accurately - choose an appropriate evaluation metric which reflects this business goal.

7.Finally, choose a model based on some evaluation metric.

As random forest is overfitting and the sensitivity is very low. So we went ahead with logistic and PCA models.

EDA – Summary

•Calls Revenue(3 columns):
•Invalid Values : Having minimum values as negatives, indicating some customers are making loss to the company. These columns are either invalid or not adding value to our prediction, can be dropped from the dataset.
•Standardise: Revenue columns can be rounded to 2 decimal places.
•Minutes of usage(60+ columns):
•Usage minutes is generally 0 except for few outliers, for below variables:
•Roaming Incoming ISD Incoming Special Incoming Others STD incoming T2F STD outgoing T2F Outgoing Others ISD Outgoing Local Outgoing T2C (Customer care calls)
•Most of the columns have outliers.
•Aggregating Columns based on Incoming and Outgoing, or Aggregating based on Each Type of Incoming Calls and Outgoing Calls and looking at the metrics will give a better understanding of the data.
•Recharge (12 Numeric + 3 Date columns)
•Data Type Conversion:
•Data in numeric columns are integers, so can be converted to int type.
•Date columns need to be converted to date type
•Data 2G And 3G(22 Columns)

EDA – Summary (cont.)

•Most of the columns have median as O and have outliers
•vbc_3g columns need column renaming as it needs month to be encoded to its number.
•Standardise: Columns can be rounded off to 2 decimal places.
•Age on Network (1 Column)
•Feature can be derived from AON column.
•Churn (Dependent Variable)
•There exists a Class Imbalance in the dataset, where actual churn customers are only ~8% of the dataset.
•Reviewing the Dropped Columns:
•More columns will be lost because of dropping missing value columns, while it can be handled to be imputed by considered 0 as missing values follow a pattern where Calls only users have blanks for Data related columns and the vice versa.
•Feature Engineering - Thoughts
•Derive no. of years the customer is using network from AON
•Derive fields to indicate the type of user the customer is: Uses Both Calls and Data, Only Calls, Only Data, Only Incoming calls, Only Outgoing calls, etc.
•Bin the customers into different segments based on Service usage, Recharge amount, Usage/Recharge pattern.
•Calls to Customer Care is a Key indicator that customer is not happy with the services, derive columns like time over call center calls

Business Insights :

Main indicator of churn:

From our model, it is clear that the factors affecting the churn are total_ic_mou_8 (Total incoming call: Minutes of usage in the action phase), total_rech_amt_diff (Total recharge amount difference), total_og_mou_8 (Total outgoing call: Minutes of usage in the action phase), arpu_8 (Average revenue per user in action phase), roam_ic_mou_8 (Roaming incoming call: Minutes of usage in the action phase), roam_og_mou_8 (Roaming outgoing call: Minutes of usage in the action phase), std_ic_mou_8(STD incoming call: Minutes of usage in the action phase), std_og_mou_8 (STD outgoing call: Minutes of usage in the action phase), av_rech_amt_data_8(average recharge amount in the action phase).

Telecom company needs to pay attention to the roaming rates. They need to provide good offers to the customers who are using services from a roaming zone. The company needs to focus on the STD and ISD rates. Perhaps, the rates are too high. Provide them with some kind of STD and ISD packages. To look into both of the issues stated above, it is recommended that the telecom company collects customer query , complaint and work on their services according to the needs of customer.

Steps to help reduce churn

- Give special discounts to customers according to their usage.
- Provide additional internet services on recharge.
- Speak to customers to fulfil their requirements.
- Lower tariffs on data usage ,a better 2G area coverage where 3G is not available.
- Expansion of 3G network where 3G is currently not available.

Thank You!