Domain Oriented Case Study - Telecom churn Case Study

Submitted By

Pritkumar Parmar

Telecom Churn Case Study

- Problem statement:-
- To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.
- In this project, we will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn. Retaining high profitable customers is the main business goal here.

Understanding and Defining Churn

- There are two main models of payment in the telecom industry - **postpaid** (customers pay a monthly/annual bill after using the services) and **prepaid** (customers pay/recharge with a certain amount in advance and then use the services).In the postpaid model, when customers want to switch to another operator, they usually inform the existing operator to terminate the services, and you directly know that this is an instance of churn.
- However, in the prepaid model, customers who want to switch to another network can simply stop using the services without any notice, and it is hard to know whether someone has actually churned or is simply not using the services temporarily (e.g. someone may be on a trip abroad for a month or two and then intend to resume using the services again).

Definition of Churn

- **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.
- **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.

High Value Churn

- In the Indian and Southeast Asian markets, approximately 80% of revenue comes from the top 20% of customers (called high-value customers). Thus, if we can reduce the churn of high-value customers, we will be able to reduce significant revenue leakage.
- In this project, you will define high-value customers based on a certain metric (mentioned later below) and predict churn only on high-value customers.

Basic Steps for Model Deployement

1) Import all the Libraries
2) Read the data SQL/Hive/csv/Excel etc
3) Data Cleaning and Data Preprocessing

- Missing Value, Outliers and coverting categorical values into continuous values.

1) EDA – Exploraterry Data Analysis / Data Visulization

- Univariate Analysis , Bivariate Analysis

1) Machine Learning Models – which colum is target variable, split the data into train-test split, select the algorithm according to the problem statement and call ML algorithm, fit on the training data and predict the values of the test data.
2) Evalute the model Performance by various parameters.

Exploratery Data Analysis

- For the Exploratery Data Analysis, used a **sweetviz** Library, so that we can do all the data visualization detailed analysis and this file will be in html format.
- Syntax of the library is as follows :
- !pip install sweetviz
- import sweetviz as sv
- sweet_report = sv.analyze(df)
- sweet_report.show_html('EDA.html')

Analysis from EDA

- Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned.
- ARPU for the not churned customers is mostly densed on the 0 to 1000.
- We can see that for the churn customers the minutes of usage for the month of August is mostly populated than the non churn customers.
- We can see that the ISD outgoing minutes of usage for the month of August for churn customers is densed approximately to zero. On the onther hand for the non churn customers it is little more than the churn customers.
- The number of mothly 3g data for August for the churn customers are very much populated aroud 1, whereas of non churn customers it spreaded accross various numbers.

Important Variables from the Model

VARIABLES OF FINAL MODEL AND ITS COEFFICIENTS

- 1) loc_ic_mou_8 :1.25
- 2) loc_ic_t2f_mou_8 : 1.20
- 3) roam_og_mou_8 : 1.03
- 4) monthly_3g_8 : 1.02
- 5) std_og_t2m_mou_8 : 1.01
- 6) og_others_8 : 1.01
- 7) ic_others_8 : 1.01
- 8) isd_og_mou_8 : 1.00

Observations

Train Data

Accuracy : 85.28%
Sensitivity : 88.87%
Specificity : 74.70%

Test Data

Accuracy : 75.00%
Sensitivity : 84.45%
Specificity : 74.64%

Recommandations

- 1. Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
- 2. Target the customers, whose outgoing others charge in July and incoming others on August are less.
- 3. Also, the customers having value based cost in the action phase increased are more likely to churn than the other customers. Hence, these customers may be a good target to provide offer.
- 4. Cutomers, whose monthly 3G recharge in August is more, are likely to be churned.

Contd.

- 5. Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
- 6. Cutomers decreasing monthly 2g usage for August are most probable to churn.
- 7. Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
- 8. roam_og_mou_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.

Thank You!