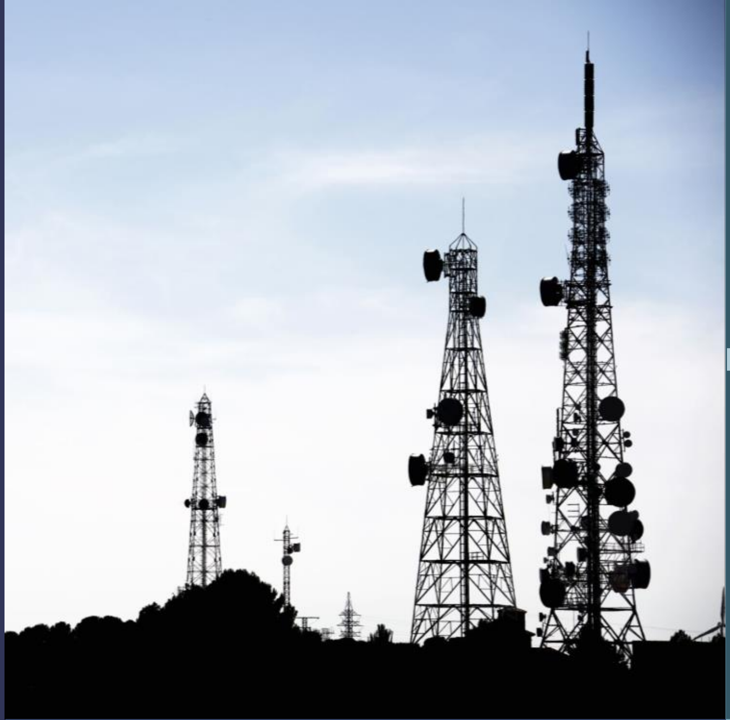
Telecom Churn Case Study



- *PRAMOD*
- *KARISHMA SINGH*
- *VIGNESHWARAN*

Problem Statement:

- The goal of this project is to predict customer churn in the telecommunications industry in India and Southeast Asia, with a focus on high-value customers. The main objective is to reduce revenue leakage by retaining high-value customers who are at risk of churning.

- To achieve this goal, you will be using customer-level data from a leading telecom firm and building predictive models to identify customers at high risk of churn. The project will use a usage-based definition of churn, and high-value customers will be defined based on a specific metric.

- This project aims to provide valuable insights into the telecommunications industry and help companies reduce customer churn by anticipating and addressing the factors that lead to it.

Business Goal:

- 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.
- By understanding the typical customer behavior during churn, you can identify the key factors that influence a customer's decision to churn. This can help you build a predictive model that accurately predicts which customers are likely to churn. Additionally, by identifying high-risk customers in the "action" phase, the company can take corrective actions to reduce the likelihood of churn. This, in turn, can help the company retain valuable customers and maintain a stable customer base.

Problem Solving Steps :

Data understanding, preparation, and feature engineering

- Understanding the Data
- Preparing the data
- Cleaning the data
- Exploratory Data Analysis

Model Building

- Preparing the data for modelling
- Building the model & Evaluating the model

- Identifying important churn indicators and business recommendation

Observation:

- We observe that almost all columns contain outliers, with most of them being due to the presence of 0.0, indicating that the service was not used. Some of the outliers, however, are actual outliers.
- We can clearly see that most of the churns are for user who had a tenure of less than 4 years
- It is evident from the data that the mobile usage minutes of the churners in the 8th month have dramatically decreased, resulting in a decline in their generated revenue
- It is noteworthy that even though the mobile usage minutes were within the range of 0-2000, the highest revenue was generated in that range, suggesting that these users utilized other services that contributed to their revenue.
- Users who recharge with high amounts seem to make fewer local outgoing calls compared to users who recharge with lower amounts.
- It is observed that even during the good phase, users with lower max recharge amounts and lower local outgoing calls have a higher tendency to churn
- The variable is not skewed but it is highly imbalanced, the number of non-churners in the dataset is around 94%
- We will handle this imbalance using SMOTE algorithm

Conclusion:

- In order to maintain our customer base and avoid loss of customers, it is imperative that we have a high recall rate. The cost of providing an offer to a customer who is unlikely to churn is lower than the cost of losing a customer and having to acquire a new one. Therefore, we need to have a high rate of correctly identifying customers who are at risk of churning, which requires a high recall rate.

- Upon comparing the performance of various models, it appears that the tuned Random Forest and AdaBoost models are the most effective, with recall rates of 97% and 95% respectively. Given the high performance of both models, we have decided to choose the Random Forest model over the AdaBoost model due to its relative simplicity.

Business Recommendation:

We can see most of the top predictors are from the action phase, as the drop in engagement is prominent in that phase

Some of the factors we noticed while performing EDA which can be clubbed with these insights are:

- Users whose maximum recharge amount is less than 200 even in the good phase, should have a tag and re-evaluated time to time as they are more likely to churn
- Users that have been with the network less than 4 years, should be monitored time to time, as from data we can see that users who have been associated with the network for less than 4 years tend to churn more
- MOU is one of the major factors, but data especially VBC if the user is not using a data pack if another factor to look out

Features
loc_og_mou_8
total_rech_num_8
monthly_3g_8
monthly_2g_8
gd_ph_loc_og_mou
gd_ph_total_rech_num
last_day_rch_amt_8
std_ic_t2t_mou_8
sachet_2g_8
aon

Thank You