Telecom Churn Case Study

Problem Statement:

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.

For many incumbent operators, *retaining high profitable customers is the number one business goal*.

To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**

In this project, you 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.

Line by line comments:

Line 1: Importing libraries

Line 4-5: Checking for null percentages and deleting variables with over 30% missing values.

Line 6-28: replacing null values with zeroes because in the dataframe all values are for same months and have the same values for said months and mostly in the same category(mou=minutes of usage) thus showing that they are meaningful missing values and the customer did not recharge.

Line 29-35: extracting high value customers and tagging churners. Churn percentage is 8.64%.

Exploratory Data Analysis:

Line 39: Recharge amount is roughly the same in the good period but declines sharply in the action period for churners.

Line 40: Average revenue per user also sharply declines amongst churners in the action period while its roughly the same in good period

Line 41-42: Both incoming and outgoing call usage is in decline for churners in the action period compared to good period while churners have fewer incoming calls in the good period compared to non-churners.

Line 43: Recharge number is roughly the same in the good period but declines sharply in the action period for churners.

Line 44: Max Recharge Amount is roughly the same in the good period but declines sharply in the action period for churners.

Line 45: Churners have a slightly lower last day recharge amount for good period but the gap widens further in the action period

Line 46-47: Churners have a significantly lower data usage for good period and the gap widens further in the action period.

Line 48: The average churner has spent significantly lower time on the network compared to non-churners.

Model Building:

Line 49-68: Decision Tree Results:-

Train Set:

Accuracy = **0.969**

Sensitivity = **0.972**

Specificity = **0.967**

Test Set:

Accuracy = **0.884**

Sensitivity = **0.668**

Specificity = **0.905**

Line 69-85: Random Forest Results:-

Train Set:
Accuracy = **0.980**

Sensitivity = **0.990**

Specificity = **0.970**

Test Set:

Accuracy = **0.923**

Sensitivity = **0.732**

Specificity = **0.941**

Line 86-162: Logistic Regression Results:-

Train Set:
Accuracy = **0.850**

Sensitivity = **0.871**

Specificity = **0.828**

Test Set:

Accuracy = **0.826**

Sensitivity = **0.797**

Specificity = **0.829**

Model Evaluation: Random Forest is an obvious upgrade over Decision Trees in all 3 departments(Accuracy, Sensitivity & Specificity) but Logistic Regression is the obvious go to choice in this scenario because while it has lower(still acceptable) accuracy compared to the other 2, it has the highest Sensitivity score on the test set which is what is more important if the focus is on identifying and preventing potential churns.

Line 163: Coefficients for different variables for churned and **Policy recommendations**:-

As expected from the chart, the least like likely to Churn all have high levels of consumption during the action period in almost every aspect like calling, recharge amount and data usage which was confirmed even in the EDA phase hence the regular high value customers who show a sudden drop in data usage/incoming-outgoing calls/recharge amount should be the first ones to be targeted for retention. Another observation that stands out from the EDA is that even in good period churners tend to have relatively low levels of data usage and incoming calls hence these 2 groups should also be the prime focus to reduce churn. And last but not the least, churners also have significantly low age on network hence high value newer customers should also be targeted to avoid potential churns.