Telecom Customer Churn Prediction

Springboard Data Science Capstone Project Proposal

Ashley Jiangyang

April 2020

The telecom industry continues to confront growing pricing pressure worldwide. Given these challenging industry dynamics, managing customer base to reduces churn should be among any senior telecom executive's highest priorities. In this project, we will use the telecom churn dataset to solve the problems of interests described below,

- What's the churn distribution for all the customer cohort (counts? Percentage for each category?)
- What information we can obtain from the dataset? Given the information, what's the relationship between that and the customer churn?
- Can we use the information to predict the customer churn?
- How can we make a recommendation for business strategies based on the insights that are leveraged from the data?

The dataset is acquired from Kaggle: Telco Customer Churn – Focused Customer Retention Programs, which contains information about,

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

In this project, we will use the telecom churn dataset to exploring the product and services dynamic develop a comprehensive view of the customer and link that view to strategy-making recommendations. Furthermore, we will build machine learning models to predict customer behavior, which can be integrated into an agile test-and-learn process to the company to customize offers for different individual microsegments, which will be helpful to improve the customer retention.

Four components will be considered to include in this project: data cleaning, exploratory data analysis, survival analysis, modeling and model comparison.