Project Proposal – Capstone Project 1

Customer churn, defined as the percentage of customers that stop using a company's products or services, is one of the most import matrices for a business, as it usually costs more to acquire new customers than it does to retain existing ones.

The objective of this project is to predict if customers are likely to stop doing business and when that event might happen.

The client is the management of the pertaining business. With the model built and data analysis performed, the business owners will be able to segment the customers based on the likelihood to churn and take appropriate actions to prevent that happening, such us launching customer retention program.

The dataset is for Telco Customer Churn, downloaded from Kaggle competition website. This dataset contains 7,043 unique observations in total, each of which represents a customer. Each column contains customer's attributes.

One of the key attributes, tenure, indicates that the time period that a customer has been with the business since the contract was signed. Since each customer has different time footings, the traditional linear machine learning algorithms don't perform well in this scenario. This is why survival analysis comes to play as the time-event correlation will be considered.

To perform survival analysis, the non-parametric method Kaplan–Meier estimator can provide a descriptive approach to consider how "tenure" affects the likelihood of customer churn by plotting survival function curve. However, semi-parametric method Cox Proportional Hazard is more powerful to predict whether a customer churn based on not only time factor but also other attributes, which can be seen from the dataset. Furthermore, some advanced techniques such as survival tree and survival random forest will also be explored.

The deliverables of this project will include code, a paper, and a slide deck.