**Springboard Data Science Career Track**

**Capstone Project - 2 Proposals**

Author: Vidyasagar Sadhanala

Date: 03-04-2020

**Proposal 1 - New York Taxi Fare Prediction**

***Objective:***

To predict the estimated taxi fare amount by using features provided by taxi cab company. The goal is to analyze the data set, perform data wrangling to manipulate and for doing explanatory data analysis (EDA), apply various machine learning algorithms to allow test data to run through these ML models to predict the expected fare price.

***Problem:***

To predict a rider’s taxi fare amount?

***Outcome:***

This estimation of rider’s fare is going to provide taxi company’s (such as Yellow Cab) who are competing against ride-hailing technology companies to provide more data centric fare estimation approach.

[***Dataset***](https://www.kaggle.com/c/new-york-city-taxi-fare-prediction/data)***:***

* Pickup Time/Date
* Pickup Latitude/Longitude
* Dropoff Latitude/Longitude
* Passenger Count
* Fare Amount (*target*)

Train set- 55m rows

Test set- 10k rows

***Evaluation Metrics:***

The evaluation metric for this problem is going to be the **RMSE** (root mean squared error), this is the difference between the predicted values and the actual values.

**Proposal 2 - Telco Customer Churn**

***Objective:***

Customer Churn measures the loss of customers and service provider companies use this metric to understand the customer retention. The objective is to predict behavior to retain customers by analyzing all relevant customer data and develop focused customer retention programs.

***Problem:***

Is the customer going churn?

***Outcome:***

The model predictions an provide the propensity of churning and gives the companies with the feature’s importance that leads the customer to churn. With the list of potential customers who are likely to churn, the marketing/retention teams can then take measure to reduce their churn probability.

[***Dataset***](https://www.kaggle.com/c/new-york-city-taxi-fare-prediction/data)***:***

* 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

Data set contains 7043 rows and 21 columns

***Evaluation Metrics:***

The evaluation metrics for this problem are going to be the **Classification Accuracy %, F1 Score, Precision, Recall.**