Laxmi Prasanna Chennadi

DATA 606

Deliverable-1 Report

**Churn Analysis of Telecom Data**

**Project Overview:**

The main goal of my project is to predict churn of telecom data which is nothing but number of people leaving current network and moving to another company. This scenario is very common and often seen in our daily life. I want to analyze the factors that effect churn which in turn would be helpful for companies in retaining their customers and also in acquiring new ones. In order to accomplish my goal, I am planning to build classification and regression models, furthermore I would perform few Machine Learning algorithms, evaluate their accuracy and then choose the best fit algorithm for the business.

**Motivation:**

Approximately 5-6 years back, during my high school, back in my country, many telecom networks entered the market, and customers started jumping from one network to another. That was the first time when question arouse in my mind. Later, after 2 years, another network came into picture which was adopted by 60% and more population of the country. This even surpassed most of the oldest and experienced companies which were in the same field since long time. These were two instances which drove me in this direction.

**DataSet:**

The dataset which will be used for this project is Telecom data which comprises of 6999 entries and 21 columns which describe the various factors that affect churn and churn column would be my target variable. It is available in Kaggle. 21 columns of dataset are, State, account length, area code, phone number, international plan, voice mail plan, number of voicemail messages, total day minutes, total day calls, day charge, evening minutes, eve calls, evening charge, night minutes, night charge, international minutes, international calls, international charge, customer service calls, churn etc.

**SOURCE:** <https://www.kaggle.com/becksddf/churn-in-telecoms-dataset>

**Methodology:**

Initially I would start with data preprocessing, i.e data cleaning, removing NaN values. Then I will perform exploratory data analysis, observe the factors influencing churn column and create visuals of the data accordingly. My next step would be applying logistic regression and Random forest classification algorithms. Later perform few other machine algorithms like KNN, PCA etc, evaluate their accuracy and choose the best algorithm which serves the business purpose. With this we will be able to predict churn, control and also improve the revenue of the company.