PROJECT PROPOSAL – PROJECT GROUP 3

ANALYSIS AND MODELLING OF CUSTOMER CHURN IN TELECOM INDUSTRY (APPLICATION BASED)

Problem Statement:

In the telecom industry, the customer churn rate is usually high. The cost of retaining an existing customer is $1/10^{th}$ of the cost of acquiring a new customer. Thus, retention of customers is far more critical than acquiring customers. To decrease the churn, the companies are looking for statistical analysis and machine learning options. The project focuses on analysis of the data for a telecom chain through Exploratory Data Analysis, Feature Extraction and Predictive Modelling. Through this project we would like to know the factors that impact churn and also understand the pattern of customers who are at high risk of churn.

Motivation:

Prepaid and post-paid are the two widely known billing options in the telecom industry. Unlike the post-paid option where customers contact service provider before switching the service, the pre-paid customers stop using the service without any prior notice. In South Asian market, 90% of the customers choose prepaid option. Thus, understanding behaviour for prepaid customers is most critical. The success of this study will give new directions in churn prediction and can improve business decision making for telecom companies.

Lifecycle of Customer Behaviour:

a) Clean/Good Phase: The service makes the customer happy and will stay with the provider.

b) Action Phase: The behaviour of the customer slowly changes from what it is in the clean phase. During this phase the customer becomes unhappy with the quality of the service and starts looking out for other network provider options.

c) Churn Phase: This is the phase where we can say that the customer has decided to leave the service.

According to our dataset, the data (features) corresponding to first 2 months describe the customer behaviour in the good phase. Furthermore, the next 2 months can be attributed to action phase and churn phase respectively.

The objective of this project is to predict the customer churn based on the data collected for 4 months (June, July, August and September respectively).

Dataset:

The source of the dataset can be found at **https://www.kaggle.com/code/gauravduttakiit/telecom-customer-churn-eda/data.**

The dataset contains 226 columns with different features such as number of calls by customer, internet usage, last billing data, etc,. The dataset contains 99999 entries.

PROJECT PLAN:

Step 1 – Understanding the Business Problem and converting it into a technical problem

Step 2 – Pre-processing of the data which involves data cleaning, removal of irrelevant data and converting the data suitable for analysis

Step 3 – Exploratory Data Analysis involves looking at various features of the data and visualizing it using plotting tools. This step is critical as it will help us understand more about data and efficient EDA will ensure a good model.

Step 4 – Feature Extraction, Selection and Model pre-processing will ensure that we acquire as much as relevant data and prepare our data for predictive modelling.

Step 5 – Building multiple classification models by using techniques like cross validation and do hyperparameter tuning

Step 6 – Model inference, understanding important predictors of customer churn and then do model comparison.