**Telco Customer Churn**

**Telco Systems is a global leader in telecommunications, with over 40 years of experience in the design and development of high-performance network communications solutions. With its advanced software and hardware solutions, Telco Systems provides a revolutionary approach to the network edge for service providers, allowing them to offer the highest levels of service innovation to customers.**

#### What is Customer Churn?

Customer churn is defined as when customers or subscribers discontinue doing business with a firm or service.

Customers in the telecom industry can choose from a variety of service providers and actively switch from one to the next. The telecommunications business has an annual churn rate of 15-25 percent in this highly competitive market.

Individualized customer retention is tough because most firms have a large number of customers and can't afford to devote much time to each of them. The costs would be too great, outweighing the additional revenue. However, if a corporation could forecast which customers are likely to leave ahead of time, it could focus customer retention efforts only on these "high risk" clients. The ultimate goal is to expand its coverage area and retrieve more customers loyalty. The core to succeed in this market lies in the customer itself.

Customer churn is a critical metric because it is much less expensive to retain existing customers than it is to acquire new customers.

**To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.**

To detect early signs of potential churn, one must first develop a holistic view of the customers and their interactions across numerous channels, including store/branch visits, product purchase histories, customer service calls, Web-based transactions, and social media interactions, to mention a few.

As a result, by addressing churn, these businesses may not only preserve their market position, but also grow and thrive. More customers they have in their network, the lower the cost of initiation and the larger the profit. As a result, the company's key focus for success is reducing client attrition and implementing effective retention strategy.

**Dataset**

The dataset used for this project was found on Kaggle. Each row represents a customer; each column contains customer’s attributes described on the column Metadata.

The data set includes 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

**Tools & Libraries**

• Python • Jupyter Notebook • Pandas • Numpy • Seaborn • Matplotlib • Plotly & Cufflinks

**Data Description**

The dataset contains the following Columns:

**CustomerID:** A unique ID that identifies each customer.

**Count:** A value used in reporting/dashboarding to sum up the number of customers in a filtered set.

**Gender:** The customer’s gender: Male, Female

**Age:** The customer’s current age, in years, at the time the fiscal quarter ended.

**Senior Citizen:** Indicates if the customer is 65 or older: Yes, No

**Married:** Indicates if the customer is married: Yes, No

**Dependents:**Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc.

**Number of Dependents:** Indicates the number of dependents that live with the customer.

**Data Cleaning**

I made the following changes and created the following variables:

• Deleted the columns URL, address and customer ID as they were not important for analysis

• Changing total charges column to float since total charges actually looks like number

• Removed Duplicate Rows

**EDA**

I looked at the different-different trends of the data and below is a few highlights of the analysis.

• I am starting our EDA by looking at the distribution of the target variable (Churn). It’s expected that the dataset is imbalanced, with less than 50% of the customers leaving the company.

• Divide the features into the following groups:

1. Demographic features
2. Services related features
3. Account information related features (categorical and numerical).

• Checking the percentage of churn for each category to understand their relationship with the target.

• Patterns in Churn Customers based on the gender

• Churn Customers based on the type of service provided

• Most profitable service types

• Features and services are most profitable

**Data Preprocessing**

#### **Splitting the data into train and test sets**

**Machine Learning Model**

Predict churn using 20% of data as test set using the following models:

* Logistic Regression;
* Random Forest;
* SVM;
* KNN;
* Decision Tree Classifier;
* Adaboost Classifier;
* Gradient Boosting Classifier;