Customer Loyalty Prediction in Telecommunications Industry: A Comparative Analysis of Machine Learning Models

By:

Deep Gori

Ikramuddin Ahmed

Shivani Latkar

Contents

- Motivation
- Objectives
- Project Workflow
- Current Progress

Motivation: Customer Churn?

- Customer Churn is defined as the percentage of customers that have stopped using a company's service.

- It is a significant issue in the telecommunication industry, with a high average churn rate of 25-30%.

- Causes: poor customer service, bad products, high prices or improperly catering the needs of different demographics .

- Telecom companies with less loyal customers have to allocate huge proportions of their budget to new customer acquisition methods such as advertisements which often drives them into loss.

- Predicting what factors increase customer loyalty and retention is always a better and cheaper alternative for these companies.

- Our project aims to address this challenge by leveraging a comprehensive range of machine learning classification models to predict customer loyalty.

Objectives

- Analyze the dataset.
- Predict customer churn and retention rate by using machine learning classification models.
- Compare the performance of different machine learning models and recommend the best one.

Project WorkFlow

- Dataset Overview

- Exploratory Data Analysis

- Data Preprocessing

- Model Building

- Model Performance

- Comparison between the models.

Current Progress

1. Dataset Overview

- The dataset was obtained from Kaggle.
- Each row represents a customer, each column contains customer's attributes described on the column Metadata. The raw data contains 7043 rows (customers) and 21 columns (features).
- The data contains the columns: 'customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges' and 'Churn'.
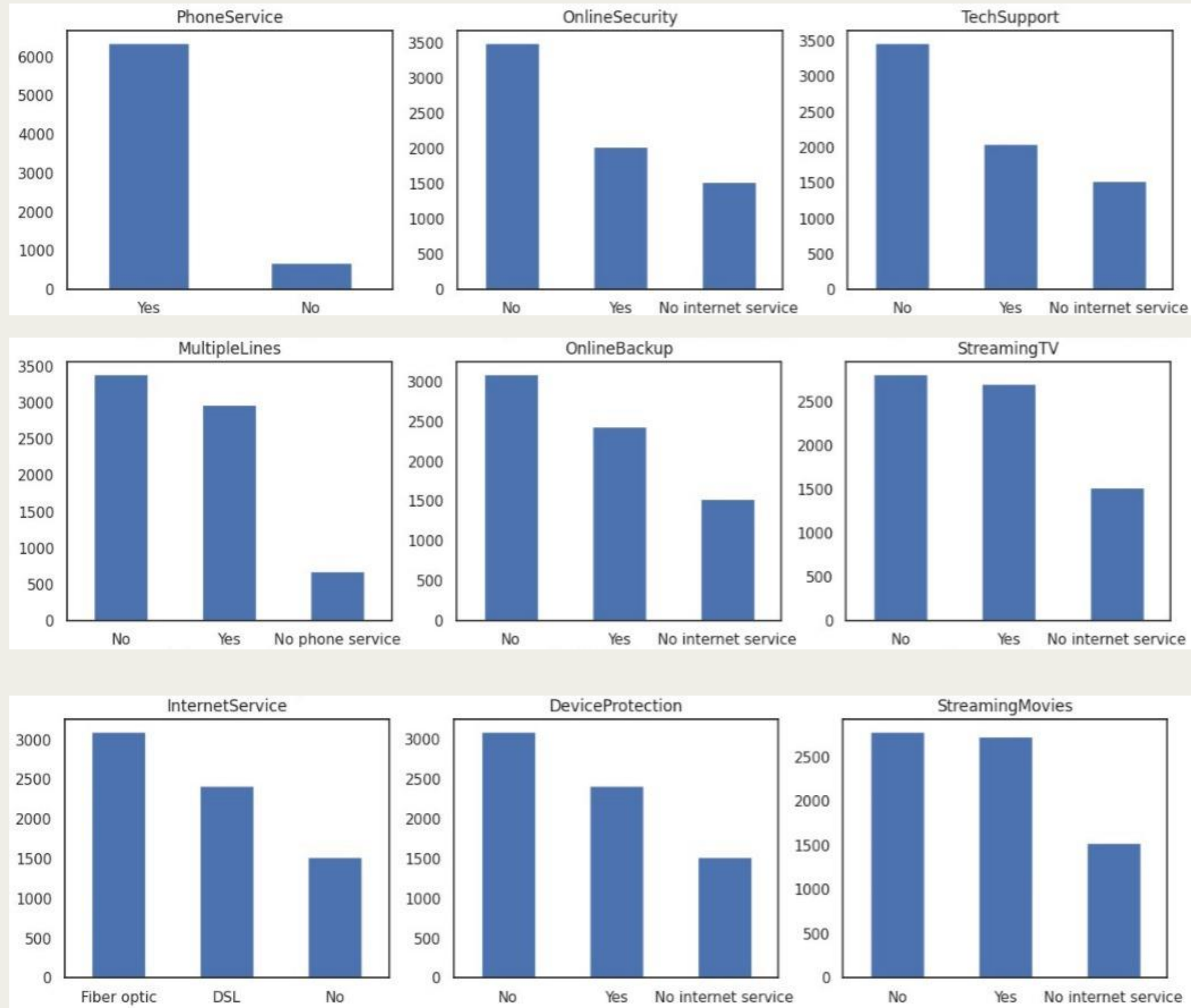
Current Progress

2. Exploratory Data Analysis
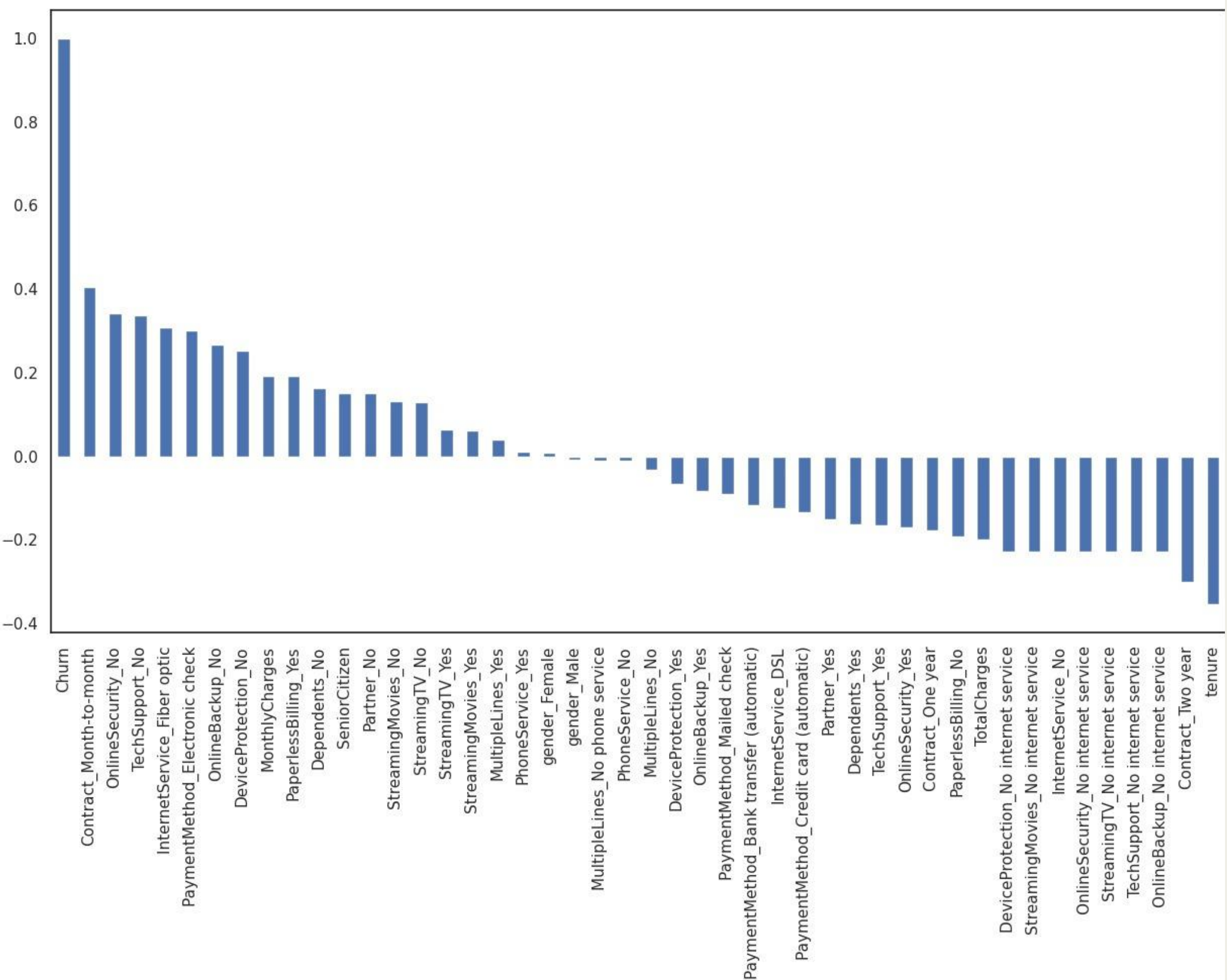
- Customer Churn in Data:
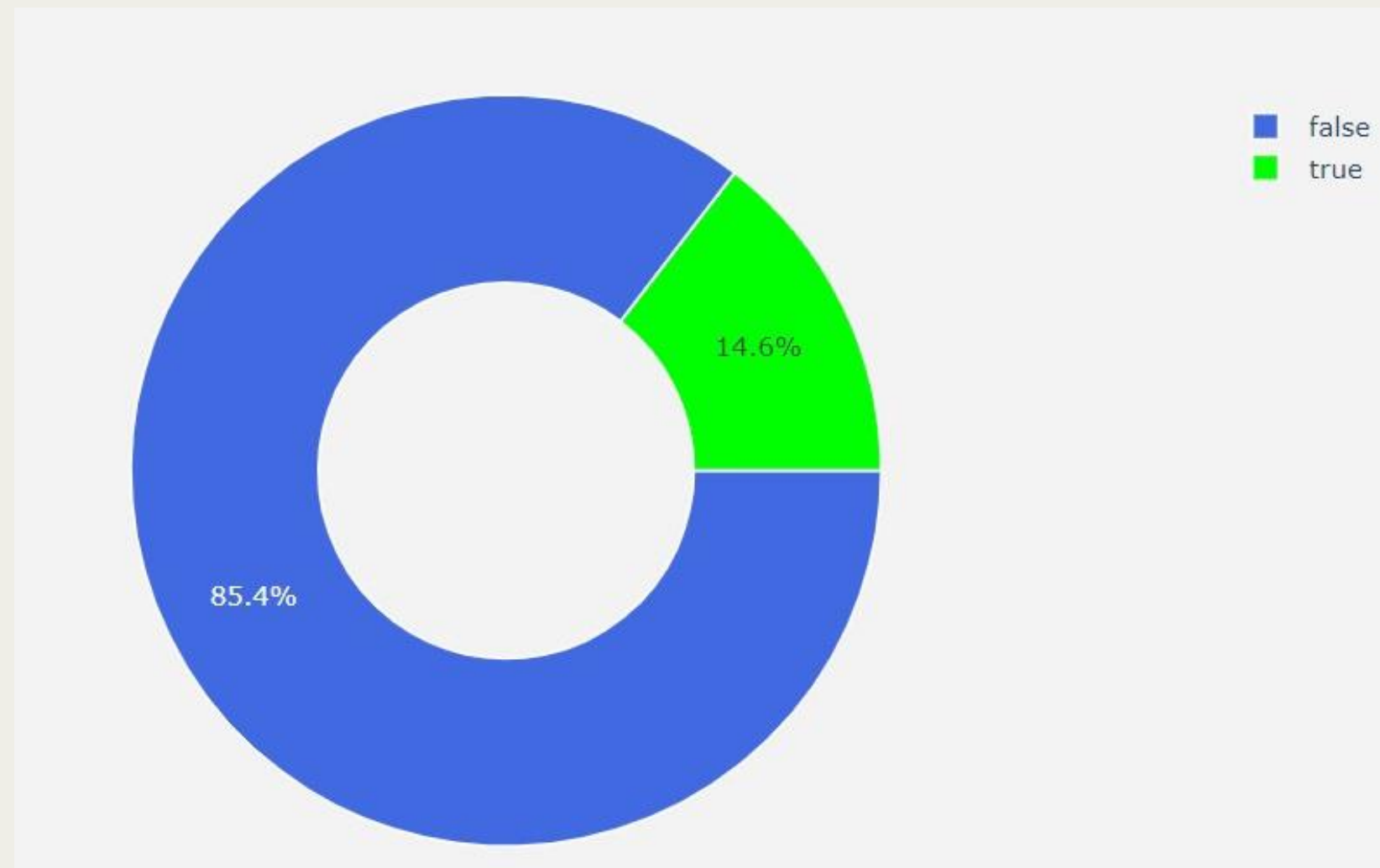
2. Exploratory Data Analysis

- Observing various services used by customers

Correlation of all features with Churn

Current Progress

3. Data Preprocessing

Data preprocessing is very important to increase the reliability and accuracy of the model as it is easy to process unambiguous data. We did the following:

- Removed rows containing null values.
- Replaced the Label's categorical values (True and False) with '1' and '0'.
- Performed One-Hot Encoding to handle all the categorical features.
- Correlated 'Churn' with other features and visualized the relationship between highly correlated features and Churn.
- Dropped unnecessary features and the features having a low correlation with Churn.

4. Model Building

We are building the following models on our dataset:

- K- Nearest Neighbor
- Logistic Regression
- Support Vector Machines
- Random Forest
- Ensemble: Soft Voting & Hard Voting Classifier

Hyperparameter Tuning:

Hyperparameter tuning is done in order to get the most optimal model for each classifier. Techniques such as grid search or random search are utilized to identify the best hyperparameters for improved model performance.

5. Model Performance

We will use the following metrics to analyze the model performance :

- Accuracy
- Precision
- Recall
- F1 score
- Area under the ROC curve

6. Comparison of Models

- We will evaluate the performance of the ensemble model and compare it to the individual models to determine if there is an improvement in prediction accuracy.
- We will analyze which model or combinations of models is the most effective for our dataset.

THANK YOU