

# Customer churn prediction model

## Project Overview

In response to SyriaTel's pressing need to address customer churn, we propose developing a robust classifier to predict customer attrition. Leveraging predictive analytics, our goal is to

uncover discernible patterns within SyriaTel's telecommunications data. By employing advanced machine learning techniques, specifically binary classification, we aim to equip SyriaTel

with a predictive model capable of identifying customers likely to churn in the near future. Our primary focus is on assisting SyriaTel in minimizing revenue loss by proactively addressing

customer retention. Through this project, we aim to unlock actionable insights that will empower SyriaTel to implement targeted strategies to mitigate customer churn, ultimately fostering

long-term business sustainability

## Research Questions

1. Which specific features in the SyriaTel dataset are the most significant predictors of customer churn, and how do these features influence the likelihood of churn?
2. How do various predictive models (Logistic Regression, Random Forest, Decision Trees) differ in their accuracy and ability to predict customer churn in the telecommunications sector, and which model provides the best balance between sensitivity (recall) and specificity?
3. To what extent does the preprocessing of the dataset, such as handling of categorical variables, normalization, and addressing class imbalance through techniques like SMOTE, affect the predictive performance of churn models?

## Goals

1. Develop machine learning models capable of forecasting customer churn through the analysis of customer characteristics.
2. Evaluate the constructed machine learning models and ascertain the most precise model for prediction.
3. The objective of the analysis is to pinpoint the particular features that exert a substantial influence on the customer churn rate at SyriaTel. By leveraging these insights, valuable recommendations can be formulated to mitigate churn rates within the company, thus enhancing customer retention

## Methodology

The project followed a standard machine learning workflow:

1. Data Loading & Cleaning
2. Exploratory Data Analysis
3. Pre-processing & Feature engineering
4. Model Development & Evaluation-Fitting various models

## 5. Model Optimization

Key libraries used include Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, and imblearn.

### **Key Findings**

Some highlights from the analysis which showed a greater impact on the customer churn rate include:

- Total Day Minutes
- Total International Minutes
- Total Day Charge
- Total Evening Charge
- Customer Service calls
- International Calls
- Total Charge ()

Random forest model performed best with a higher level of accuracy with highest AUC score

### **Repository Contents**

- Jupyter Notebook containing full project code and analysis.
- `syriatel` - original dataset.
- `README.md`