SYRIATEL PREDICTIVE ANALYSIS OF CUSTOMER CHURN

Winny Chemusian

Project overview

SyriaTel, a telecommunications company bases in Damascus Syria, encounters a notable obstacle in curtailing customer churn.
Customer retention is crucial for sustained success, as high churn rates can significantly impact revenue and profitability.
Identifying common indicators such as usage patterns, billing history, and customer service interactions is crucial for predicting and preventing churn.

OBJECTIVES

The primary objectives of this project are as follows:

- To Build a classification model to predict customer churn for SyriaTel.
- To Identify the key factors influencing customer churn.
- To Provide insights and recommendations to SyriaTel for effective churn management.

Data

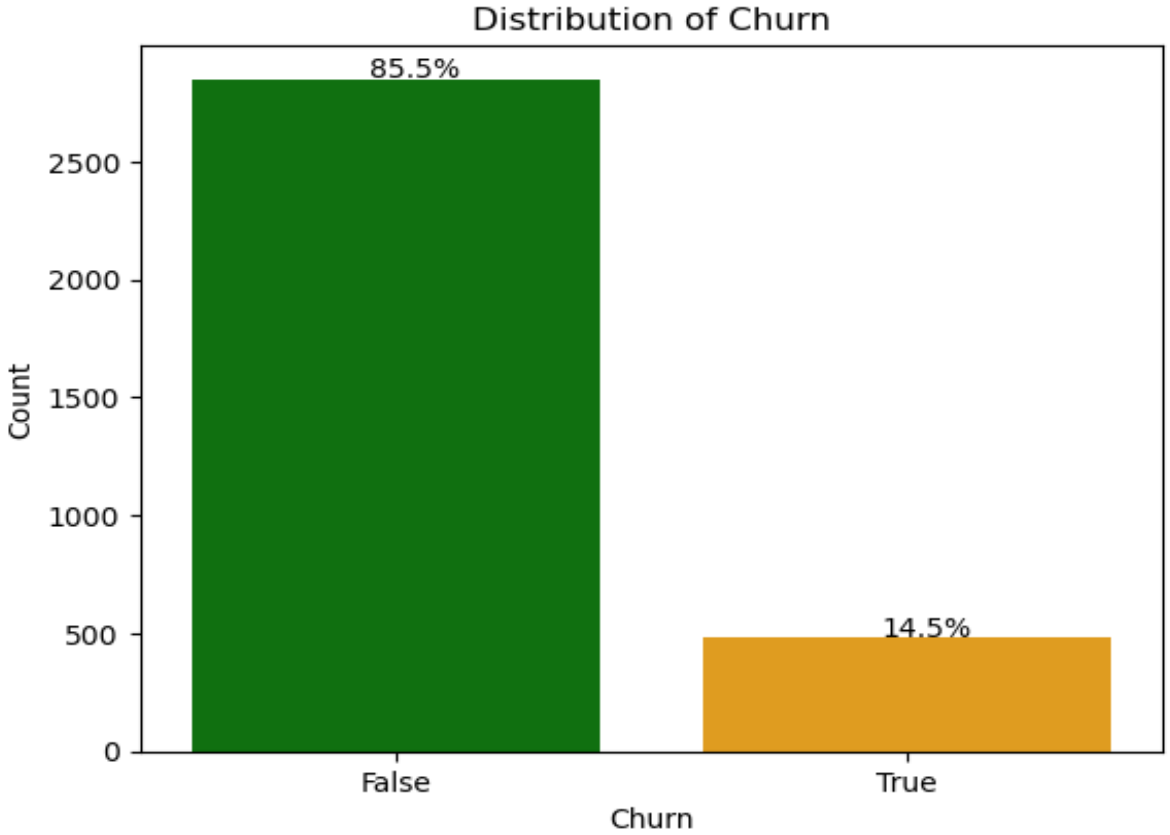
- The data utilized for this project has been sourced from Kaggle.
- The dataset contains 3333 entries and 21 columns.

Methodology

The processes undertaken in this project are:

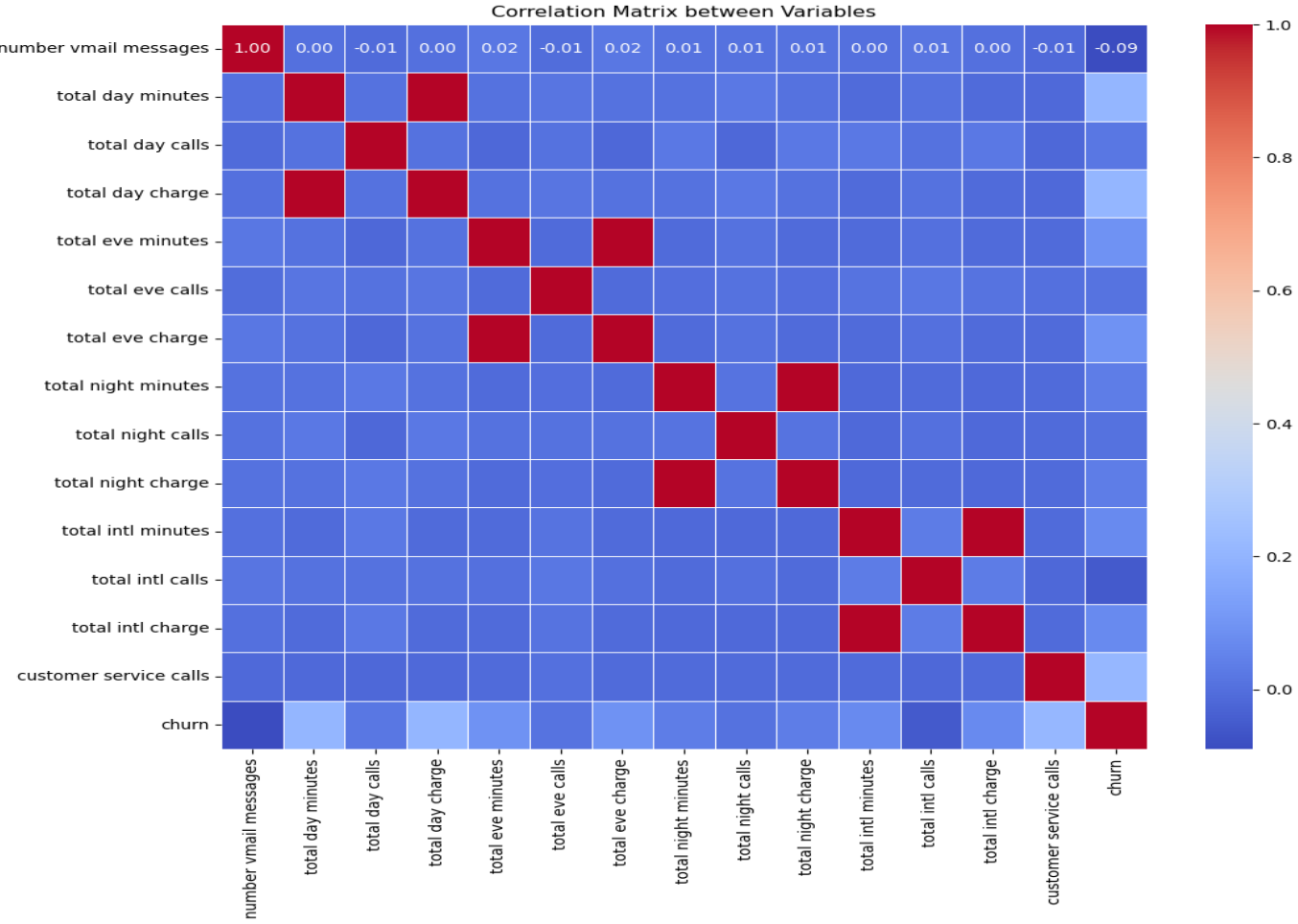
- Data Understanding
- Data Cleaning
- Exploratory Data Analysis
- Data Preparation
- Modelling
- Evaluation
- Conclusion

Univariate Analysis: Churn Distribution



From the distribution as shown in "Distribution of churn" graph, their is an uneven distribution of observations with 85.5% of the data belonging to the False class while 14.5% belonging to the true class.

Bivariate Analysis



- Due to high multicollinearity between each other, we dropped the following columns:

- total day minutes
- total eve minutes
- Total intl minutes
- total night minutes

Modelling and model Evaluation

Performed 5 classification models in this project as outlined in the table below with their scores,with the logistic regression as the baseline model.

Model	Precision	Recall	F1 Score	Accuracy	ROC AUC Score
logistic regression	0.6839	0.7369	0.7095	0.7061	0.7866
Tuned decision Tree	0.8560	0.7928	0.8232	0.8342	0.8968
RandomForestClassifier	0.9369	0.9099	0.9232	0.9263	0.9775
Gradient Boosting	0.9403	0.9369	0.9386	0.9404	0.9839
K-Nearest Neighbour	0.7825	0.9532	0.8595	0.8482	0.9408

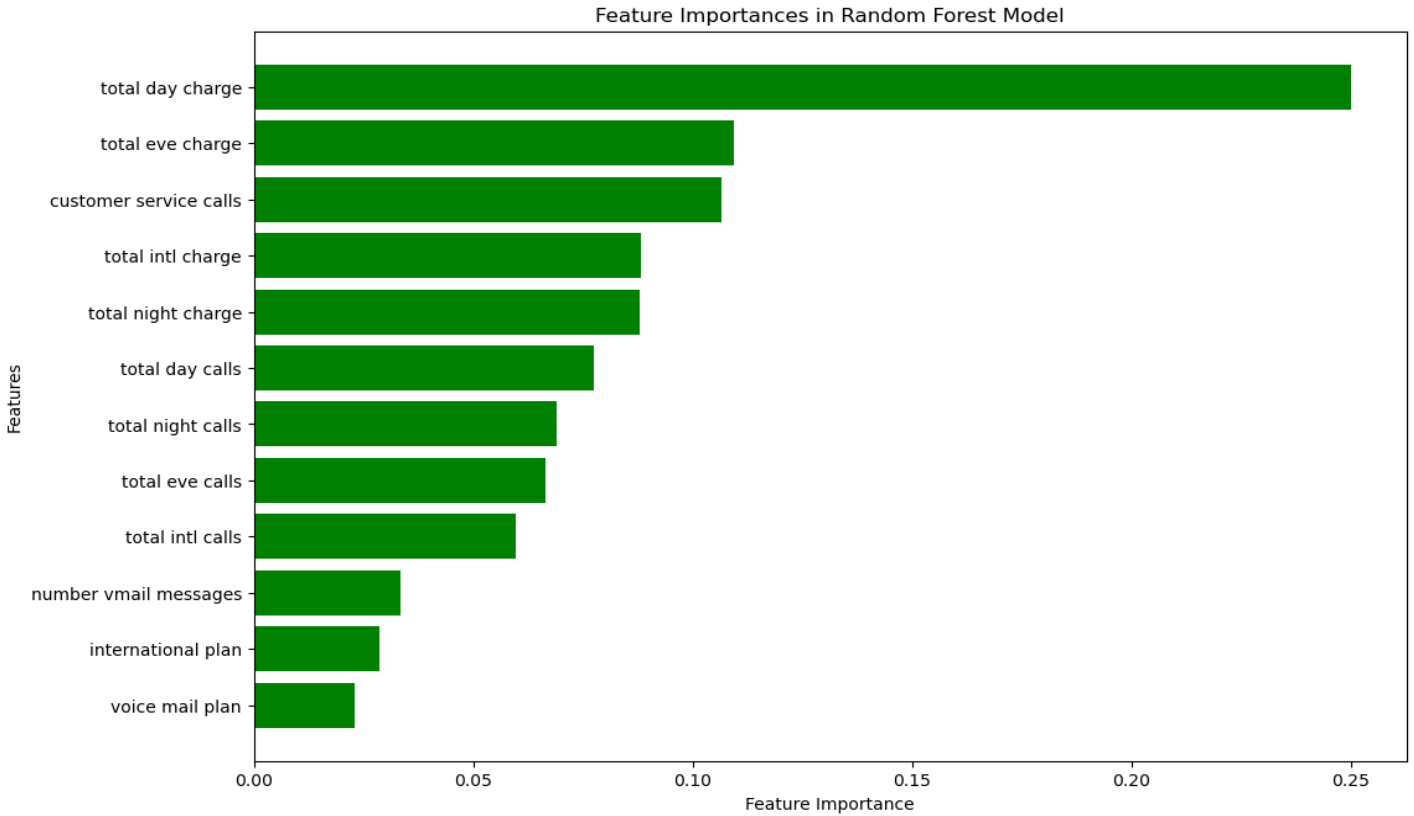
MODEL Results interpretation;

- 1.Logistic Regression: Provides a baseline with reasonable performance but lower compared to other models.
- 2.Tuned Decision Tree: Improved performance over logistic regression with higher precision and accuracy.
- 3.RandomForestClassifier: High precision and recall, indicating strong performance in predicting churn.
- 4.Gradient Boosting: Best overall performance across all metrics, with the highest ROC AUC score.
- 5.K-Nearest Neighbour: High recall but lower precision compared to RandomForest and Gradient Boosting.

Best Model: The Gradient Boosting model

- GB model is the best performer with an ROC AUC score of 0.9839, indicating the highest accuracy in distinguishing between churners and non-churners. These scores suggest that ensemble methods like Gradient Boosting and Random Forest are highly effective for the customer churn prediction task, offering superior performance compared to individual models like logistic regression and decision trees.

factors Affecting Customer Churn



The most important features for predicting churn are:

- Total day charge
- Customer service call
- Total Evening charge
- Total International charge
- Total night charge

Recommendations

Based on our findings, these are the recommendations in order to reduce customer churn:

•Focus retention strategies on high-usage customers and those with frequent customer service interactions. e.g offering discounts and incentives.

•Investigate the low adoption of international and voice mail plans to understand customer needs and improve these offerings e.g offering more affordable international plans, or by making it easier for customers to sign up for international plans .

•Provide proactive support to customers making frequent customer service calls to improve their experience and satisfaction.

•Continuously monitor and analyze usage patterns to detect early signs of potential churn and act accordingly.

EMAIL: chemusianw@gmail.com.
GITHUB: Chemusian.
LINKEDIN: Winny Chemusian.