SyriaTel Customer Churn Prediction

Prepared by: Felix Mwendwa

Project Overview

Customer churn is a critical issue for SyriaTel, directly affecting profitability and long-term business sustainability. This project aims to develop a machine learning-based churn prediction model to identify customers who are likely to leave. Therefore, through the analysis of historical customer data and predictive analysis, retention strategies can be applied to encourage customers to stay, thereby reducing churn rates and improving customer loyalty.

Business Understanding

1. Build a predictive model for customer churn.

2. Identify high-risk customers.

3. Recommend retention strategies.

Data Understanding

SyriaTel is a telecommunication company based in Syria, whose dataset of 20 features and 3,333 rows

I used the dataset to create a predictive model to help lower their churn rate of 15% (Customers who decided to end their subscription).

Data Preparation

I immediately checked for missing values which there had not been any in the entire dataset.

I also checked for duplicated values or rows that can be eliminated under the circumstance that one may have accidently slipped through, but there were no duplicated values.

I also made some modifications to the data without changing its meaning.

Distribution of Customer Churn
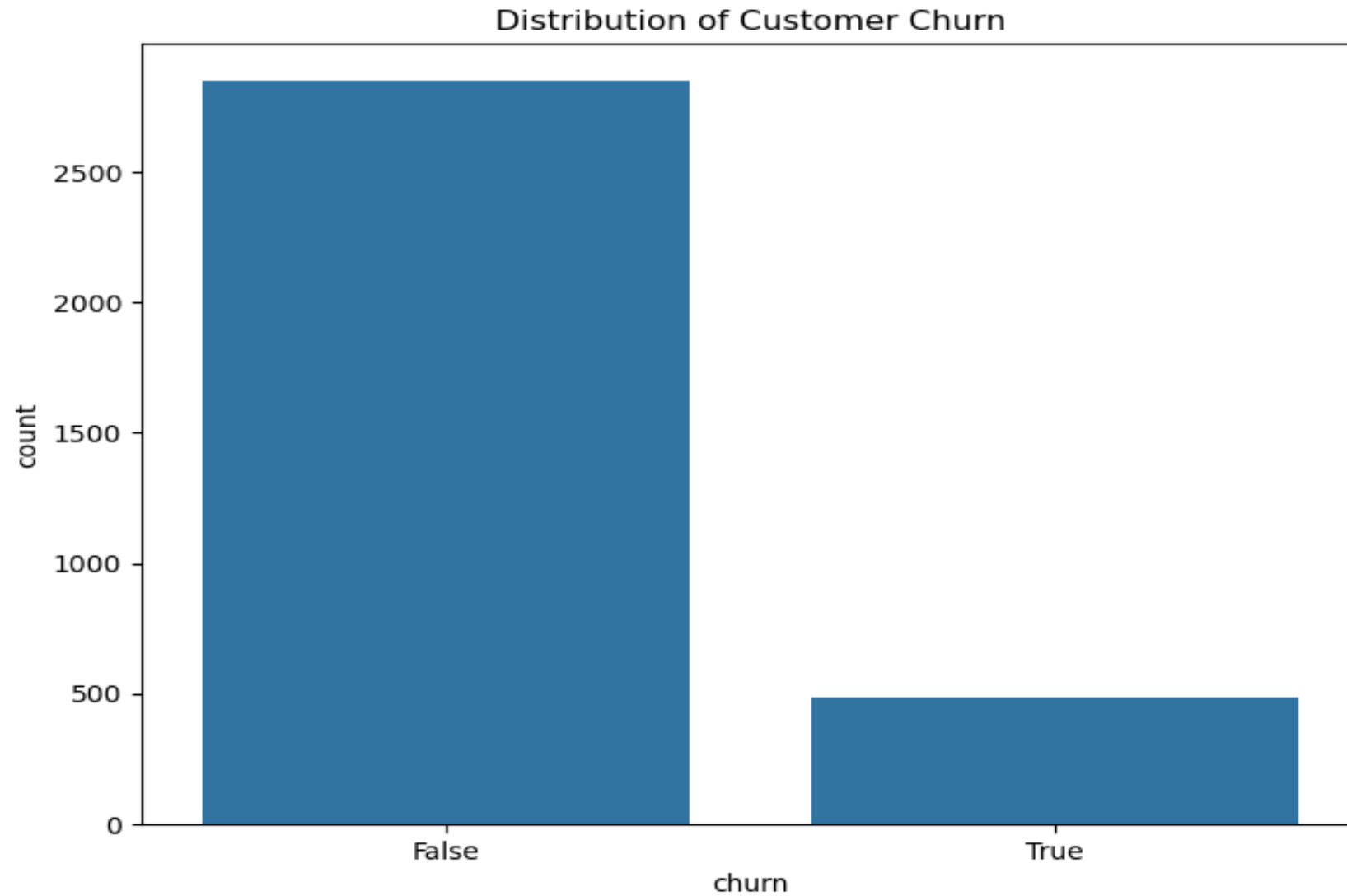
Observation

There is a significant larger number of customers classified as non-churners (false), compared to churners (true).

Modeling

Customer churn prediction is a binary classification problem, where the goal is to predict whether a customer will churn (1) or stay (0).

Classification is crucial for solving the churn problem because it helps businesses predict, prevent, and manage customer effectively, leading to higher retention rates, better resource allocation, and improved profitability.

Modeling

```
Optimized Random Forest Performance:
          precision    recall  f1-score   support

       0       0.97      0.99      0.98       570
       1       0.91      0.81      0.86        97

accuracy                          0.96       667
macro avg       0.94      0.90      0.92       667
weighted avg    0.96      0.96      0.96       667

Accuracy Score: 96.1 %
```

Confusion Matrix - RandomForestClassifier(random_state=42)
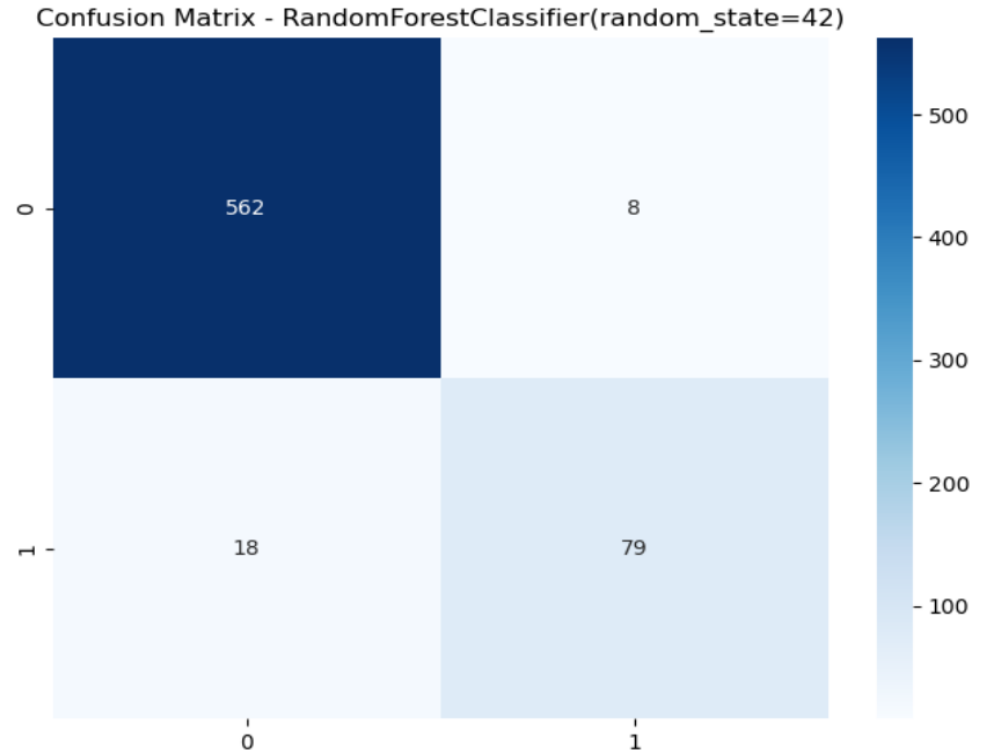
The True Negatives are 562, meaning 562 customers were correctly predicted not to churn.

The True Positives are 79, meaning 79 customers were correctly predicted to churn.

The False Negatives are 8, meaning 8 customers actually churned but were predicted not to.

The False Positives are 18, meaning 18 customers were wrongly predicted to churn but they did not actually churn.
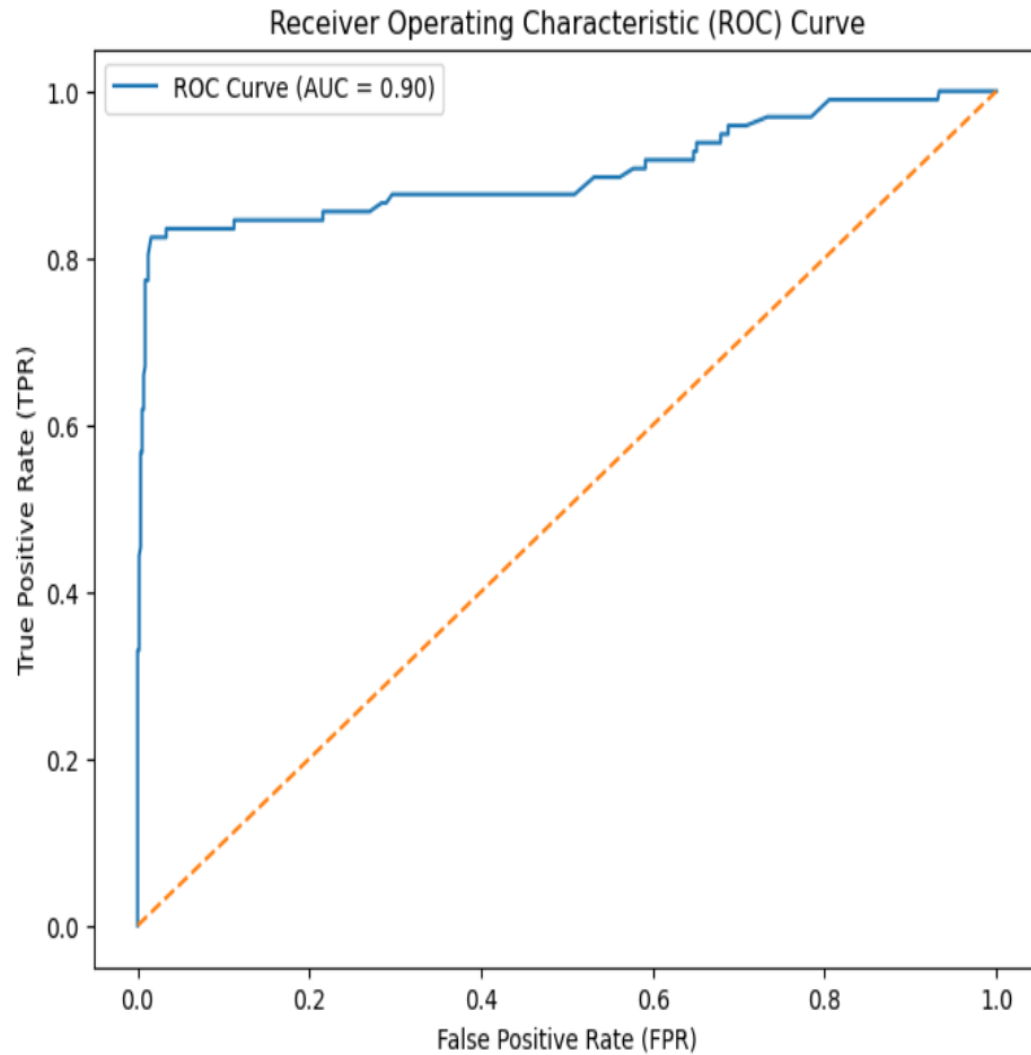
AUC Score : 0.90

- An AUC of 0.90 suggests strong predictive power (90% chance that a randomly chosen churner is ranked higher than a non-churner)

- The model's ROC curve is well above the diagonal, indicating good classification ability.

- The steep initial rise is ideal as the model captures a high TPR (Recall) early while keeping FPR low.
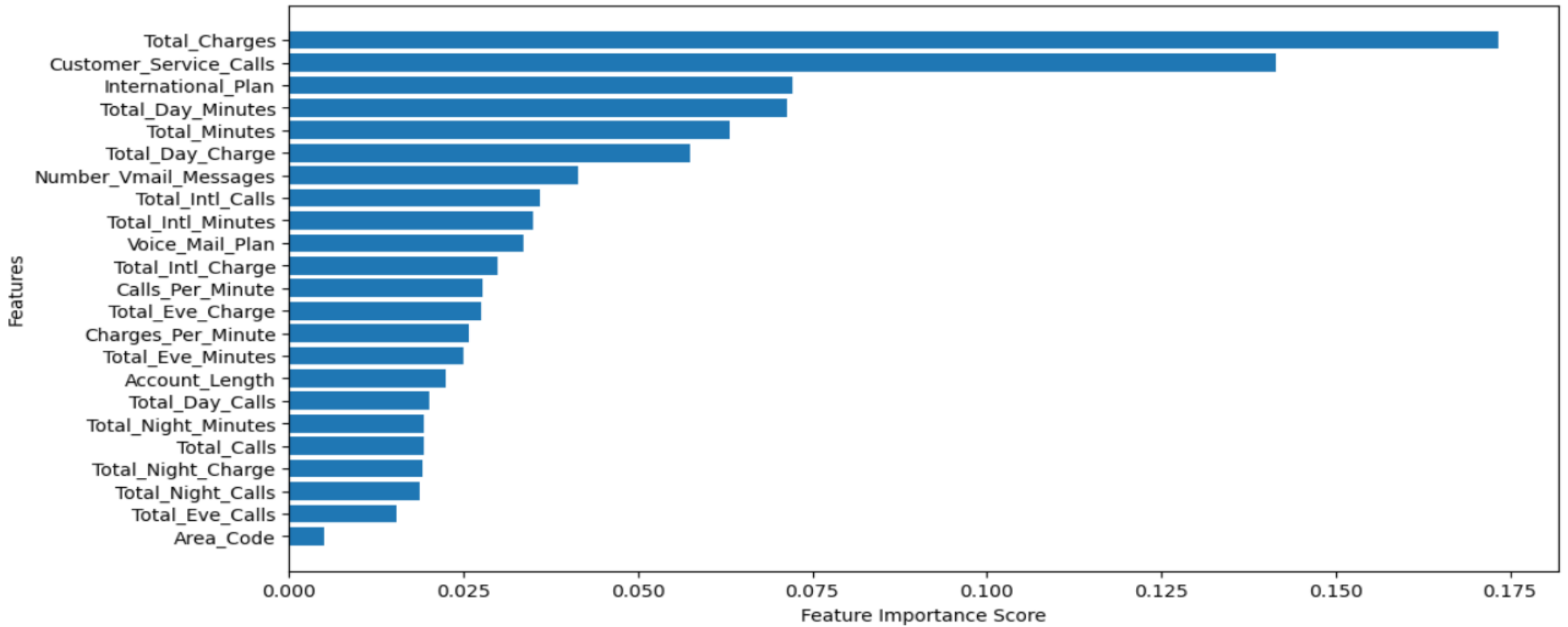
Observation

The most important features are:

• Total_Charges

• Customer_Service_Calls

Evaluation

1. The preferred classification model for predicting customer churn is the Random Forest model with an adjusted prediction threshold.

2. The initial Random Forest model had high accuracy but a recall of 0.74 for churners, indicating that many actual churners were misclassified.

3. Using GridSearchCV, we optimized parameters like max_depth, n_estimators, and min_samples_split, improving overall performance but not significantly increasing recall for churners.

4. Adjusting the classification threshold led to a recall increase from 0.74 to 0.81, meaning more actual churners were correctly identified. Precision remained strong at 0.91, and overall accuracy was 96.1%.

5. The AUC score of 0.90 indicates a well-performing model with a good trade-off between sensitivity and specificity.

6. The most important features influencing churn were:

• Total Charges (higher charges increase churn likelihood)

• Customer Service Calls (frequent calls correlate with churn).

• International Plan (subscribers to this plan are more likely to churn).

Recommendations

- SyriaTel should implement the optimized Random Forest model with threshold adjustment, as it balances recall and precision effectively.

- The company should target high-risk customers (e.g., those with high charges and frequent customer service interactions) with personalized offers and better customer support.

- SyriaTel should improve customer service quality to reduce dissatisfaction and churn risk.

- The company should offer incentives to customers on the International Plan to increase retention.

- Further improvements can be made by exploring alternative models such as XGBoost.

Thank You!

Feel free to reach out if you need any further information