SyriaTel Customer Churn Prediction

Business Problem

SyriaTel has been experiencing a high churn rate leading to significant losses. This project aims to create a predictive model that accurately identifies customers at risk of churning. The objective is to decrease customer attrition, retain a higher number of customers, and ultimately reduce financial losses while improving overall retention rates and business strategies.

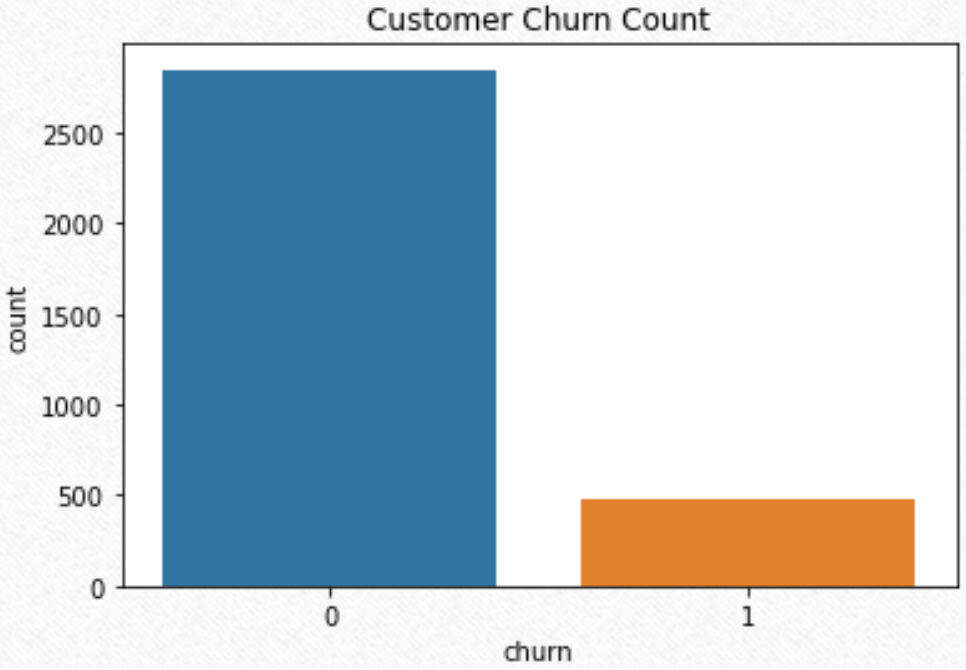
Objectives

1. Identify Key Predictors of Customer Churn: Determine which features most significantly influence customer churn at SyriaTel.
2. Provide Actionable Insights for Customer Retention by translating the findings from the predictive models into actionable recommendations for the management and marketing teams at SyriaTel.
3. Build and compare different machine learning models that is, logistic regression and decision trees to predict customer churn with high accuracy.

Data Understanding

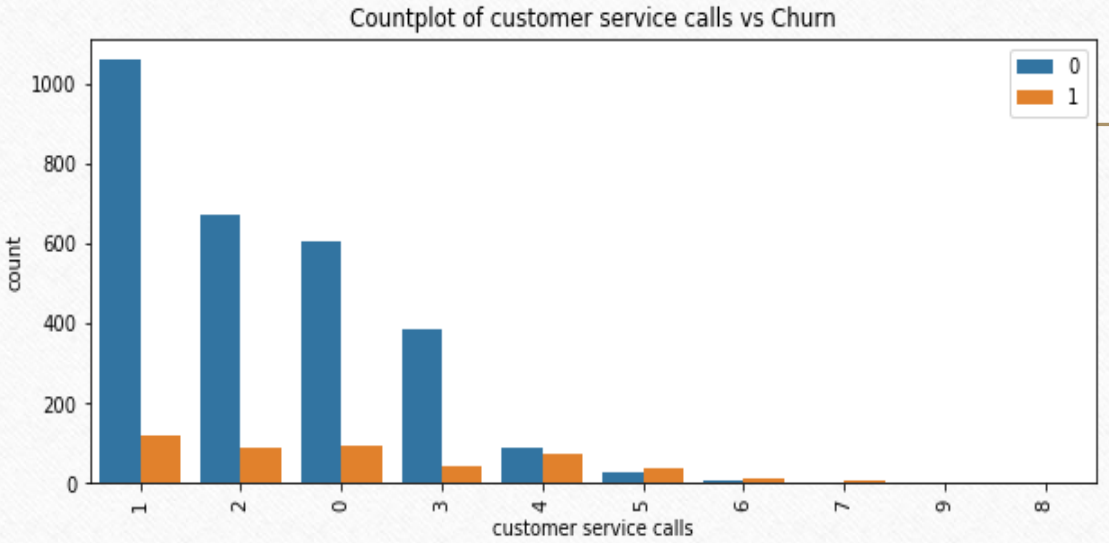
The dataset contains information on 3,333 customers from SyriaTel, with 21 features including demographic data, service usage, and customer interaction metrics. The target variable is churn, indicating whether a customer has left the company.

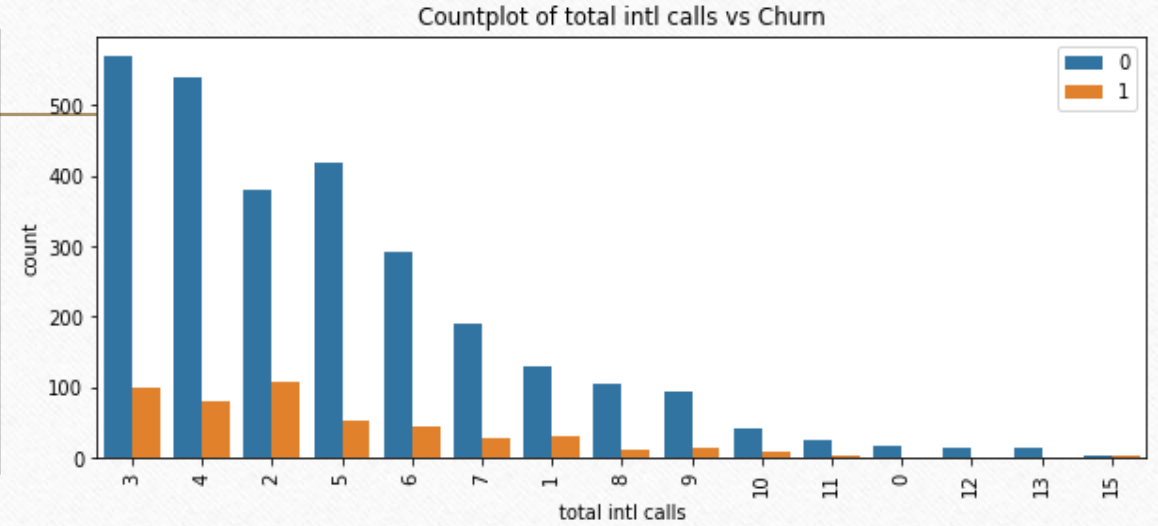
Exploratory Data Analysis



We observed a class imbalance in the churn data, with significantly more customers not churning compared to those who did.

Exploratory Data Analysis





The analysis revealed that the frequency of customer service calls and the subscription status to international plans were significant factors influencing customer churn.

Feature Engineering

Feature engineering involved creating combined features, ratios, and flags to enhance the predictive power of the model. Redundant features were dropped to simplify the model.

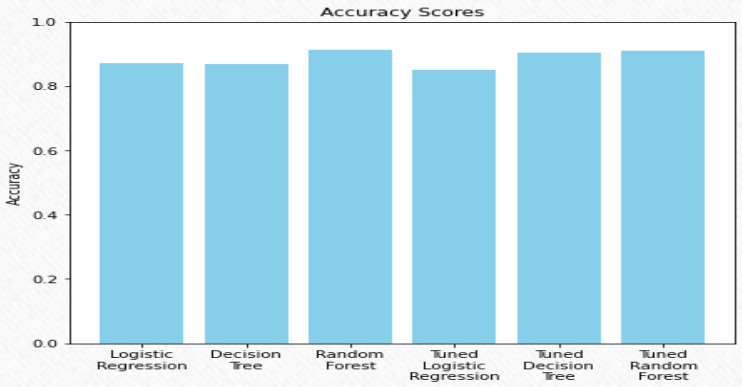
Modeling

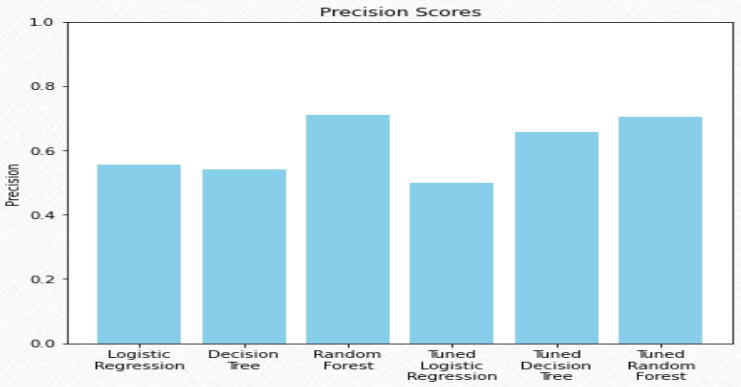
Three models were developed:

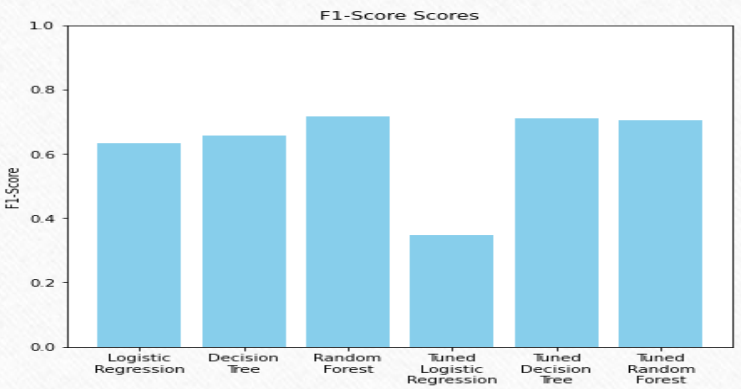
- Logistic Regression,
- Decision Trees,
- Random Forests.

The models were optimized with hyperparameter tuning. Each model was evaluated using metrics such as precision, recall, accuracy, and F1-score.

Model Evaluation

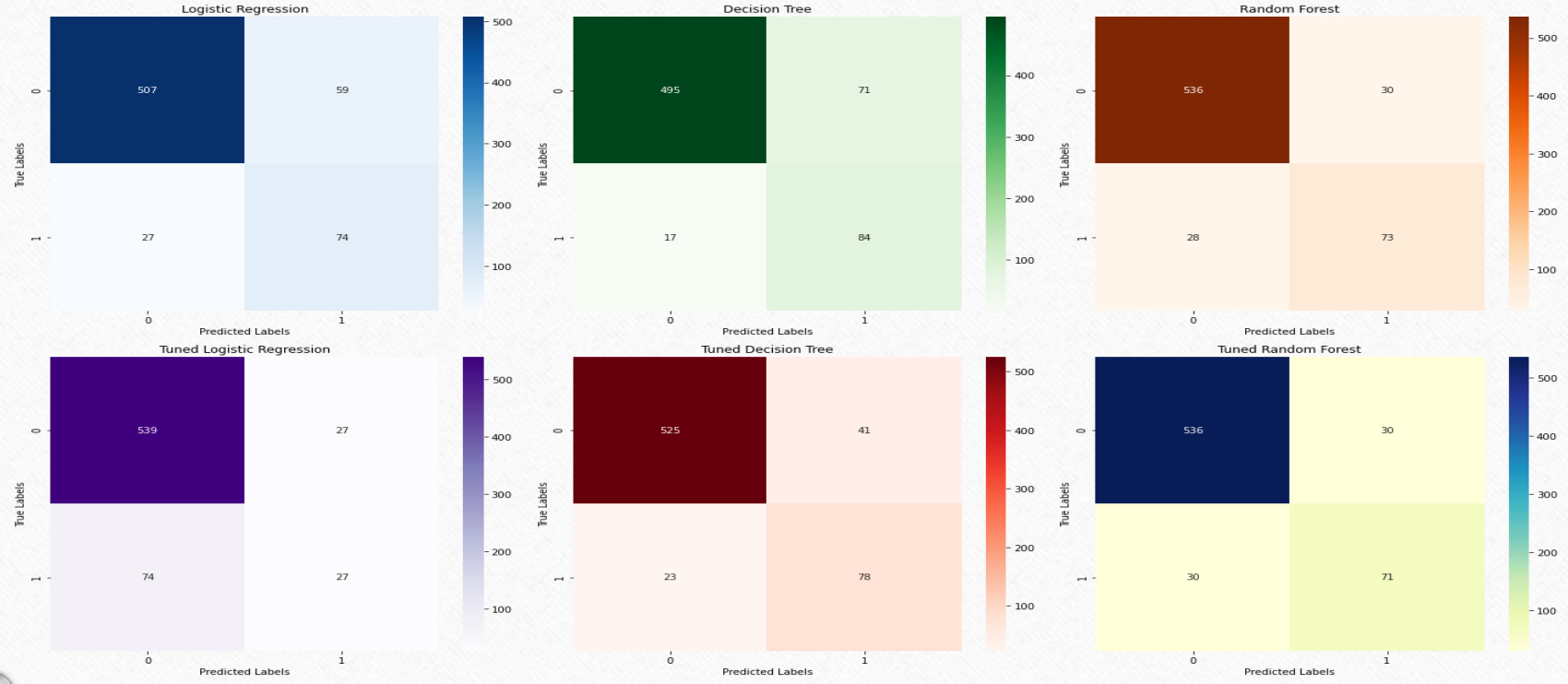






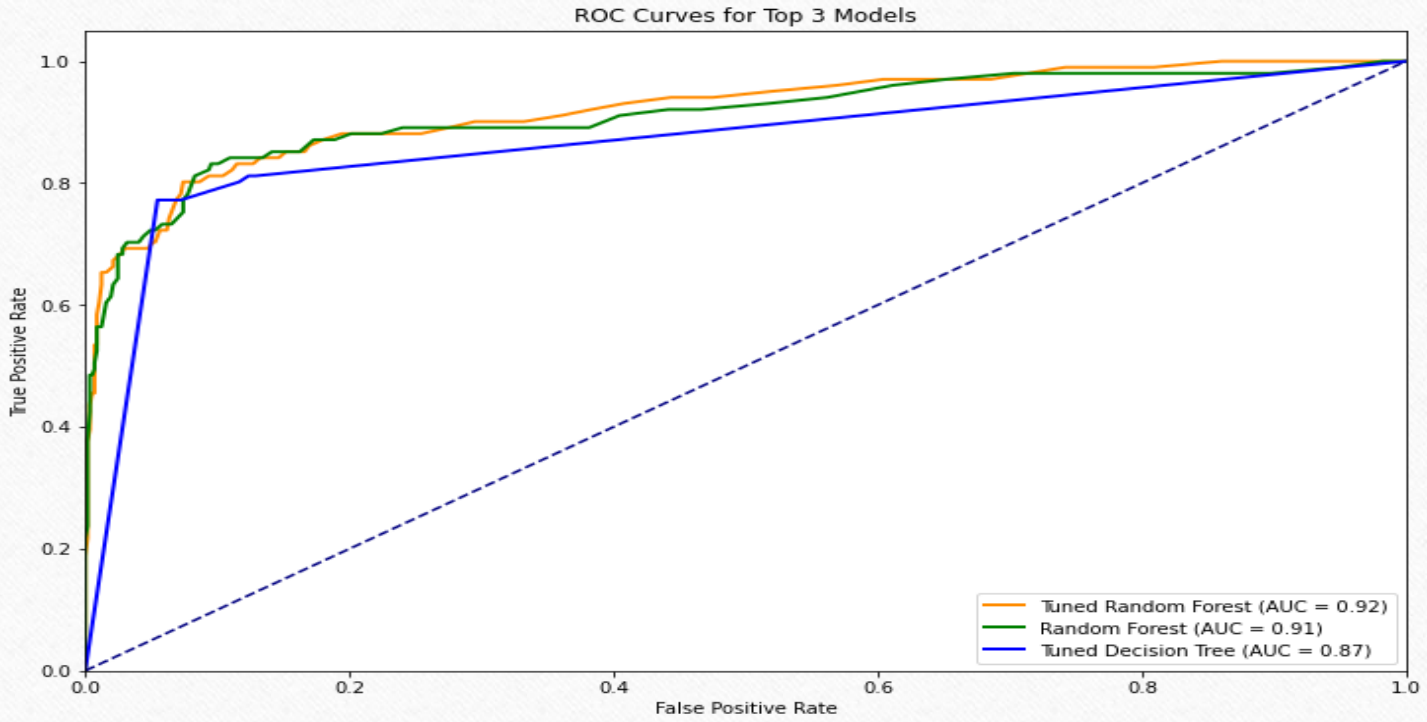
The tuned decision tree model offers a good balance between precision, recall, accuracy, and F1-score on the testing set, indicating that it generalizes better to new data compared to the other two models.

Model Evaluation



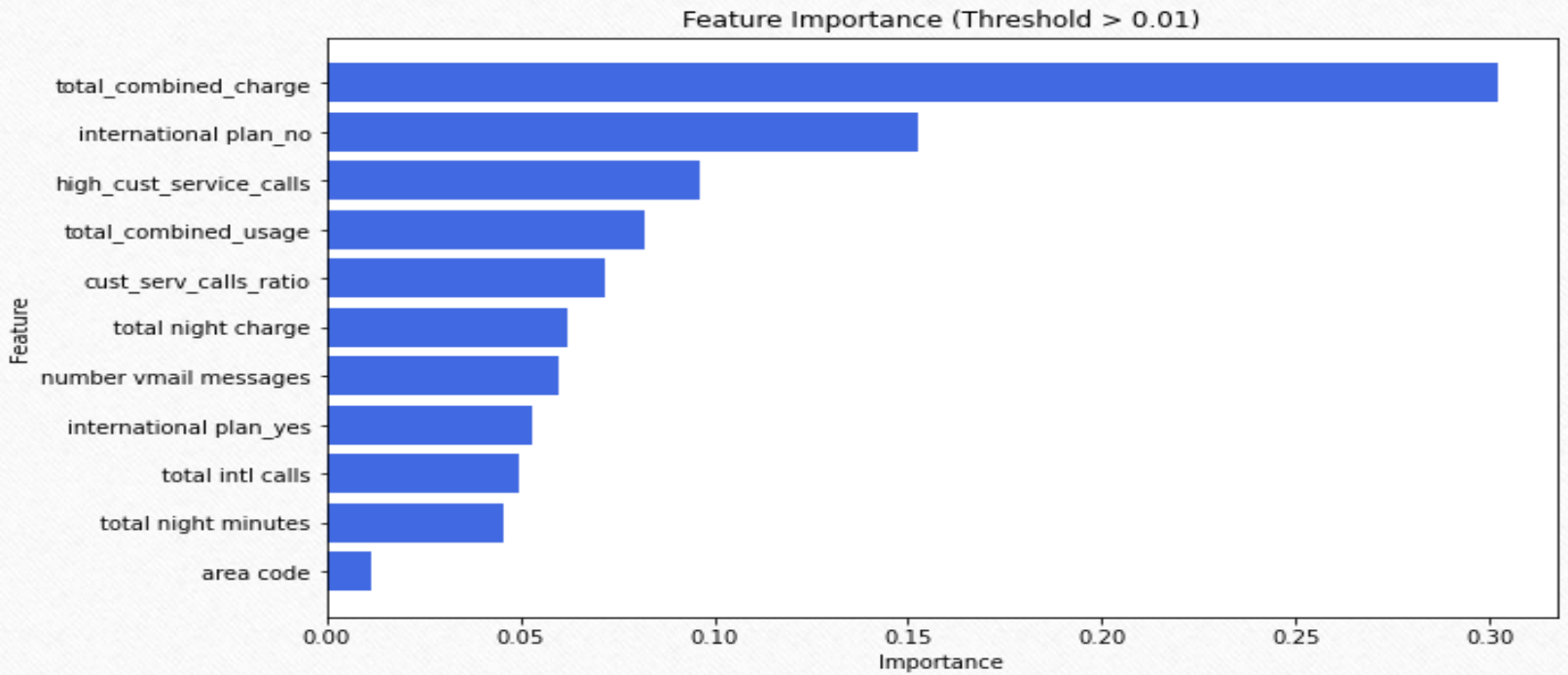
The decision tree model indicates effective identification while minimizing false negatives

Model Evaluation



The tuned random forest model performs well, the significant overfitting and slightly lower generalization performance makes it less desirable. The tuned logistic regression model is not recommended due to its poor recall and F1-score on the testing data.

Feature Importance



The Decision Tree model identified the most important features influencing churn, including:

1. Total Combined Charge
2. International Plan (No)
High Customer Service Calls
3. Total Combined Usage

Key Findings

1. Total Combined Charge is the most critical factor in predicting customer churn.
2. Customers without an international plan are more likely to churn.
3. High customer service call frequency is a strong indicator of potential churn.

Recommendation

- Focus retention efforts on customers with high total charges, as they are at the greatest risk of churn. Consider offering discounts, personalized plans, or loyalty rewards to increase perceived value.
- Improve Customer Service: Reduce churn by improving the customer service experience, ensuring that issues are resolved efficiently and that customers feel their concerns are being addressed.
- Given the importance of the international plan features, SyriaTel should consider re-evaluating these offerings to ensure they meet customer needs and are priced competitively.

Thank you