SyriaTel Customer Churn Analysis

Presented by Victoria Baikozhoeva

11/02/23

Overview

Customer churn is one of the most important metrics for a growing business to evaluate as it is much less expensive to retain existing customers than it is to acquire new customers. Customers in the telecom industry can choose from a variety of service providers and actively switch from one to the next.The technical progress and the increasing number of operators has raised the level of competition. Companies are working hard to survive in this competitive market depending on multiple strategies. This becomes a problem, as Telecom companies usually incur huge costs to attract subscribers.S

Problem Statement

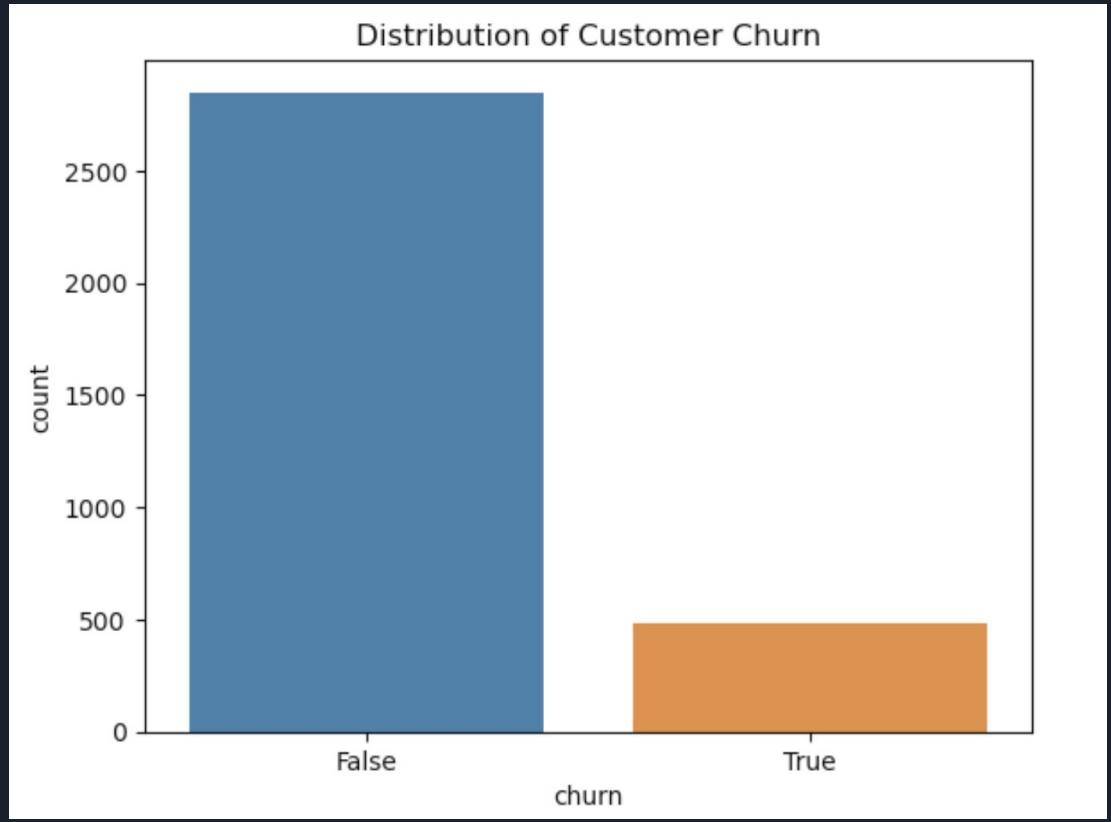
Across the telecommunication industry, customer churn is one of the most important concerns that directly affect a telecommunication company's business. Hence, companies that are able to successfully predict customer churn will be able to allocate capital and resources more efficiently and thereby improve profitability.

This project analyzes customer churn (customers leaving the provider) data from the telecommunications provider SyriaTel. The main objective of this project is to identify what type of customers were churning and develop a model that could predict whether a customer is likely to churn.

EXPLORATORY DATA ANALYSIS

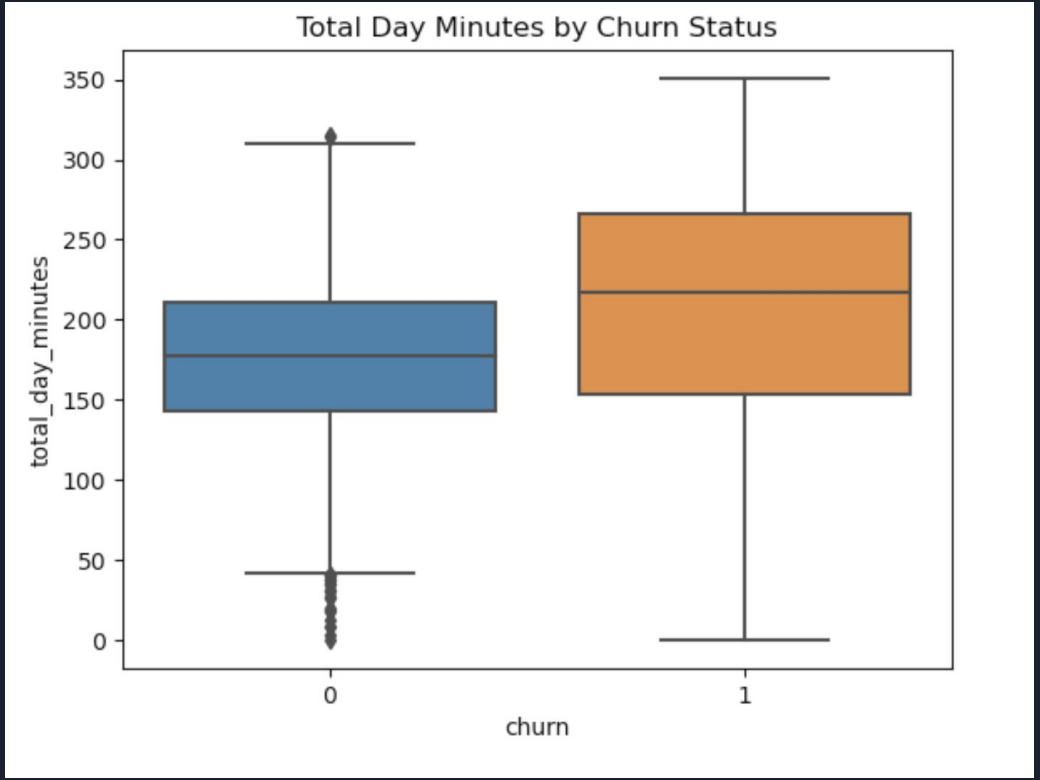
Churned Customers Count Plot

EDA found that our target's distribution is fairly imbalanced, with only approximately 14% of the customers in the dataset belonging to the 'churn' class.



Box plot of total day minutes by customer churn

From this graphs we see that feature ‘Total Minutes Per Day’ has noticeable difference between customers who churned vs those who did not churn.

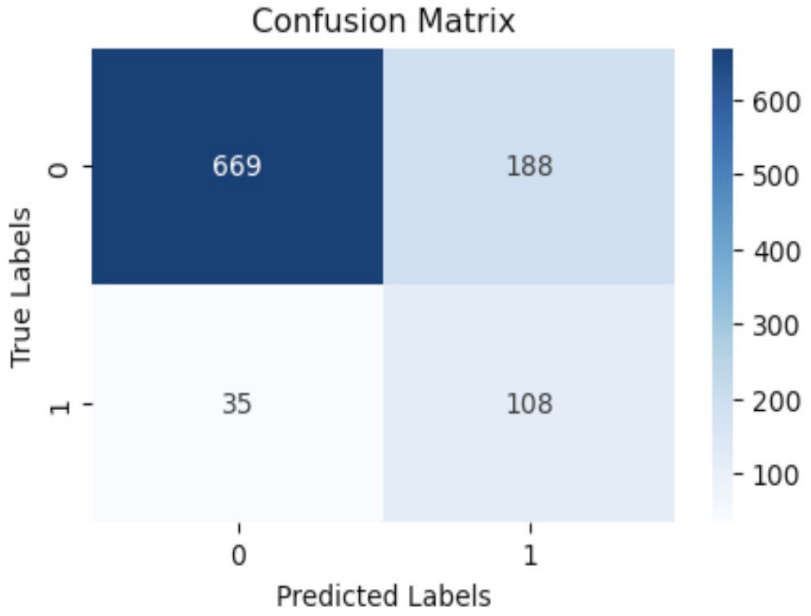


MODELING

M1 Logistic Regression Classifier

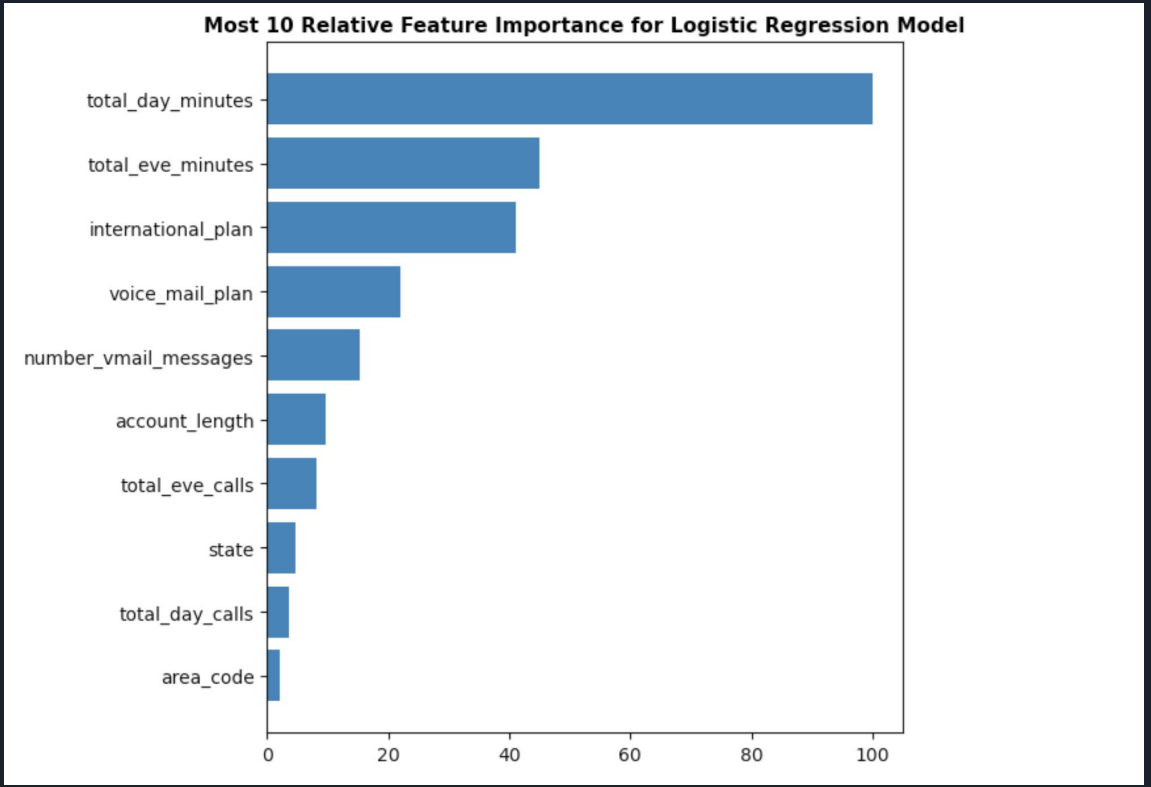
Model accuracy is almost 78%, which isn't bad. F1 score is only 49% which means the test will only be accurate half the times it is ran.

```
------------ LOGISTIC REGRESSION CLASSIFIER MODEL RESULTS------------
Accuracy score for testing set:  0.777
F1 score for testing set:  0.49203
Recall score for testing set:  0.75524
Precision score for testing set:  0.36486
```



Linear Regression Model Feature Importances

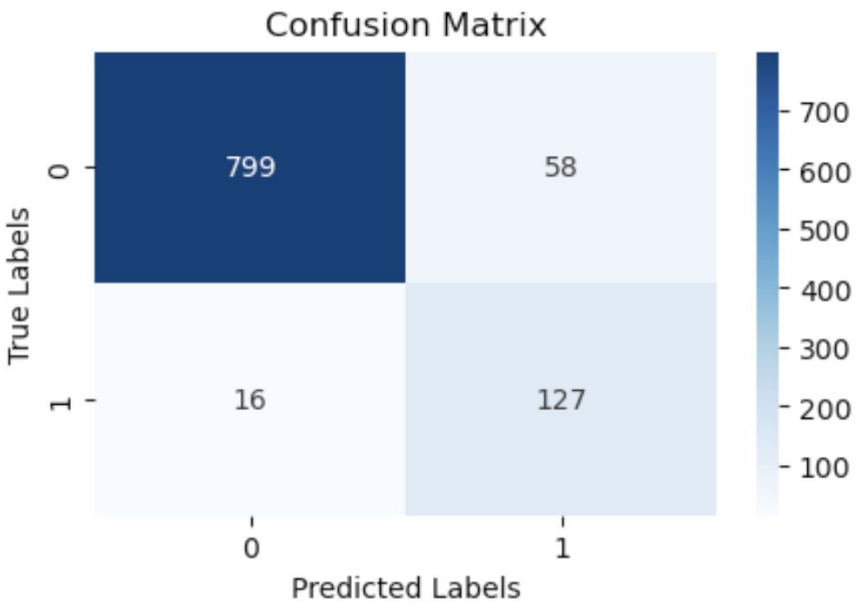
According to the logistic regression classifier model, total day minutes, total evening minutes and international plan are the top three important features



M2 Decision Tree Classifier

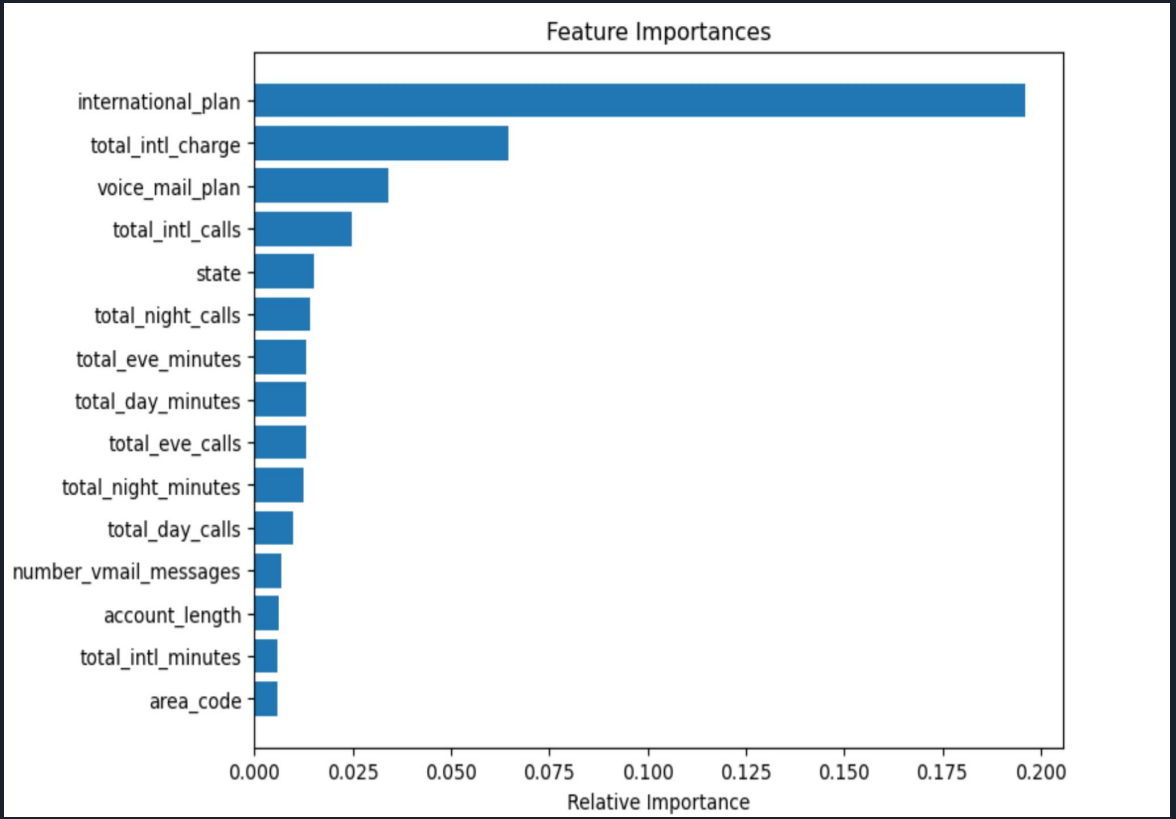
The model accuracy is almost 93% and F1 score is 78%, recall and precision scores are much better than model 1

```
------------- DECISION TREE CLASSIFIER MODEL RESULTS -------------
Accuracy score for testing set:  0.926
F1 score for testing set:  0.77439
Recall score for testing set:  0.88811
Precision score for testing set:  0.68649
```



Decision Tree Classifier Feature Importances

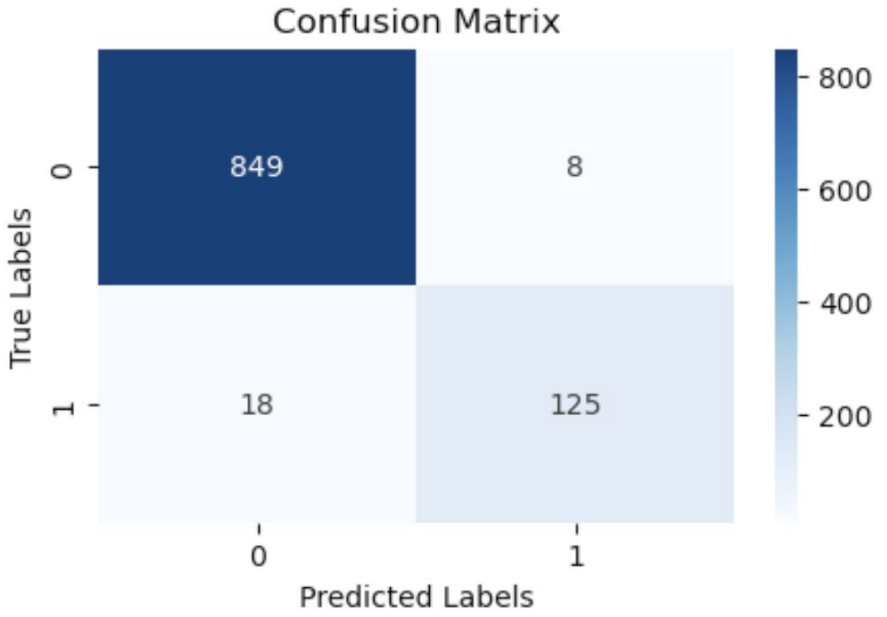
According to the decision tree classifier, International plan, total International minutes and number main messages are the three most important for the model.



M3 Random Forest Classifier

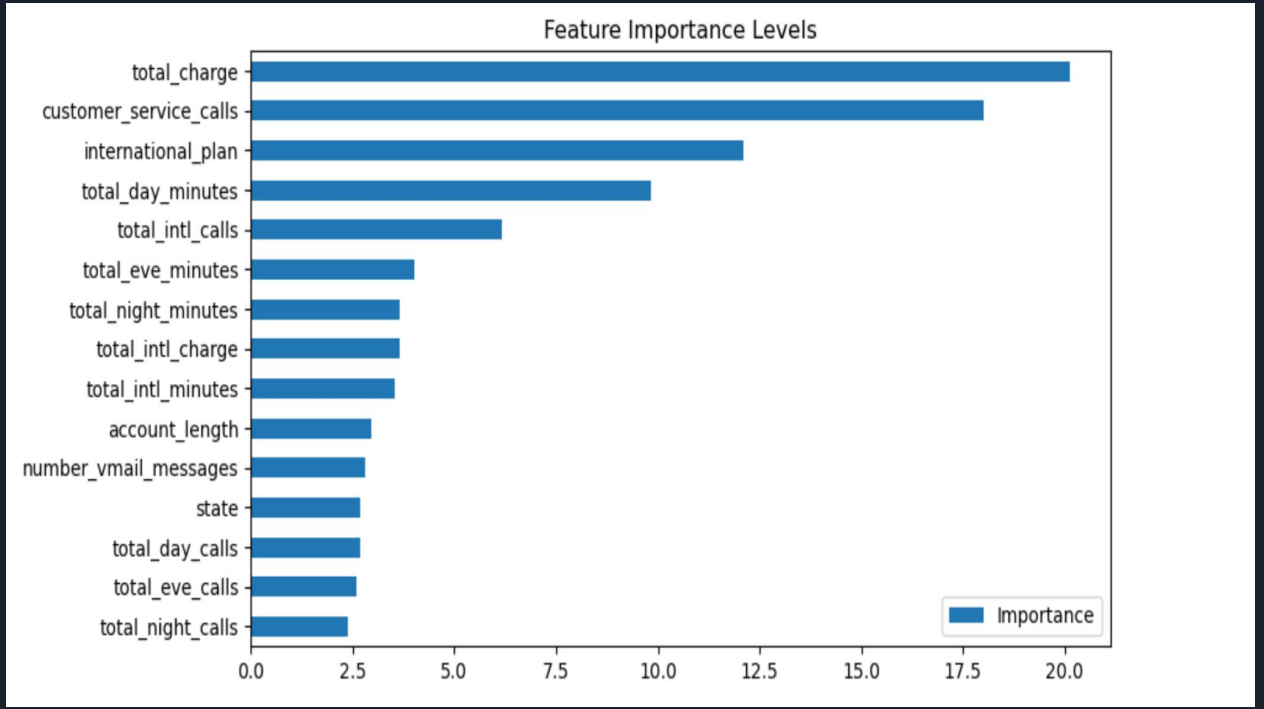
The model accuracy is almost 98% and F1 score is 91%, recall and precision scores are much better than model 2.

```
------------- RANDOM FOREST MODEL RESULTS -------------
Accuracy score for testing set:  0.974
F1 score for testing set:  0.9058
Recall score for testing set:  0.87413
Precision score for testing set:  0.93985
```



Random Forest Classifier Feature Importances

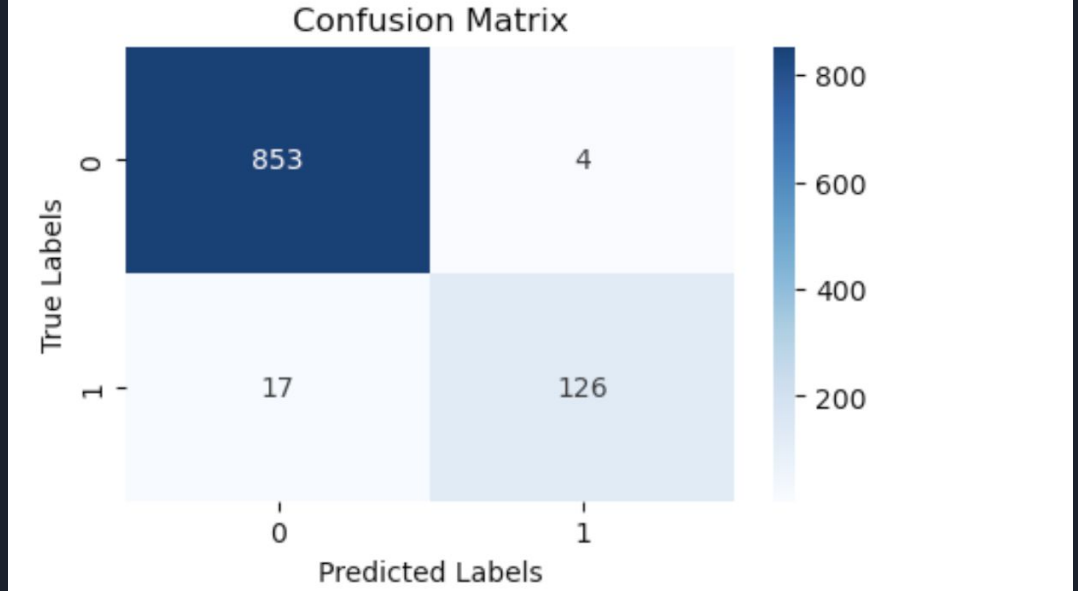
According to the Random Forest Classifier, Total Charge, Customer Service Calls and International Plan have the highest impact on the model.



M4 Hyperparameter Tuning of Random Forest Classifier

Cross validation GridSearchCV hyperparameter tuning technique is used.

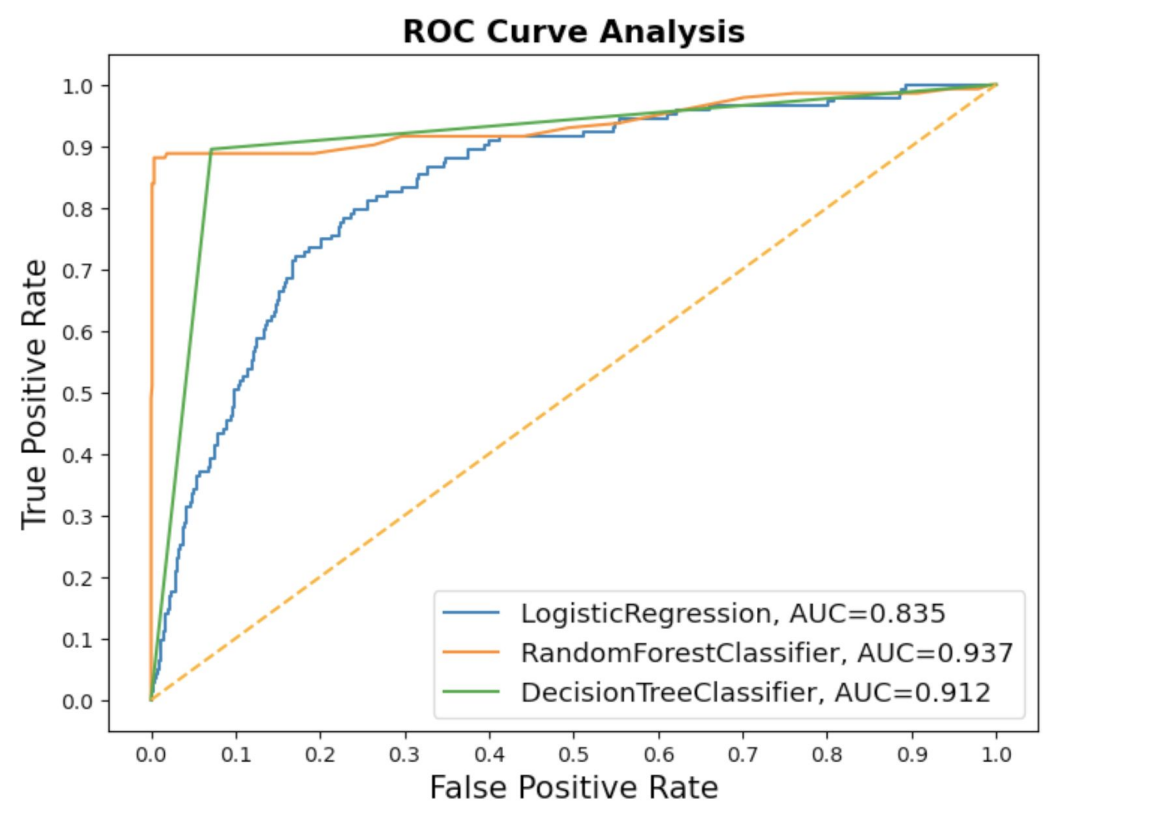
```
------------  HYPERPARAMETER TUNED RANDOM FOREST MODEL RESULTS-----
Accuracy score for testing set:  0.979
F1 score for testing set:  0.92308
Recall score for testing set:  0.88112
Precision score for testing set:  0.96923
```



Models Comparison

The ROC curve plots the true positive rate (sensitivity or recall) against the false positive rate (1 - specificity) for all possible threshold values.

The best performing models will have a curve that hugs the upper left of the graph, which is tuned random forest classifier in this case.



Modeling Summary

- The accuracy score of 0.979 means that the model correctly predicted the outcome for 97.9% of the observations in the test set.
- The F1 score of 0.92308 is the harmonic mean of precision and recall and is used to balance the precision and recall. The higher the F1 score, the better the balance between precision and recall. In this case, the F1 score of 0.92308 is a good indication that the model has a good balance between precision and recall.
- The recall score of 0.88112 means that the model correctly predicted 88.112% of the positive outcomes in the test set. This score is important when the positive outcome is of great interest and we want to minimize false negatives.
- The precision score of 0.96923 means that when the model predicts positive, it is correct 96.923% of the time. This score is important when false positives are costly and we want to minimize them.

Recommendations

- Rate Assessment

It is essential to evaluate the current charging rates and identify any potential areas for improvement, as customers seem to be dissatisfied with high charges, resulting in increased likelihood of churn.

- Customer Service Evaluation

To improve customer experience, it is crucial to assess the quality of customer service offered by the company. With over 50% of customers making more than three service calls churning, additional training for customer service staff and the creation of an internal forum to document common customer issues should be considered.

- International Plan Analysis

To stay ahead of the competition, it is crucial to investigate the viability of offering an international plan to improve retention and customer satisfaction. A thorough market analysis should be conducted to assess the competitiveness of pricing against other providers, with relevant improvements made to the international plan based on the results.

- Voicemail Plan Promotion

The low number of customers subscribed to the voicemail plan suggests that some customers may be unaware of the option, and therefore promoting this plan may help to reduce churn.