
SyriaTel Customer Churn
Prediction
SIEMENS
FEUER
NOTRUF
Brandon Muraya
March 10 2024

Overview

A telecommunication company, SyriaTel wants to know whether a customer will stop doing business with them. This understanding is crucial to telecommunication companies as predicting customer churn can directly increase their revenue or market share.

With this information, companies can implement targeted strategies to retain customers, such as offering incentives, personalized promotions, or improved customer service. This proactive approach can significantly impact customer retention rates and, consequently, the overall financial health of the company.

Data Exploration

After going through the dataset provided by SyriaTel, there were 3333 total entries from the 21 columns available.

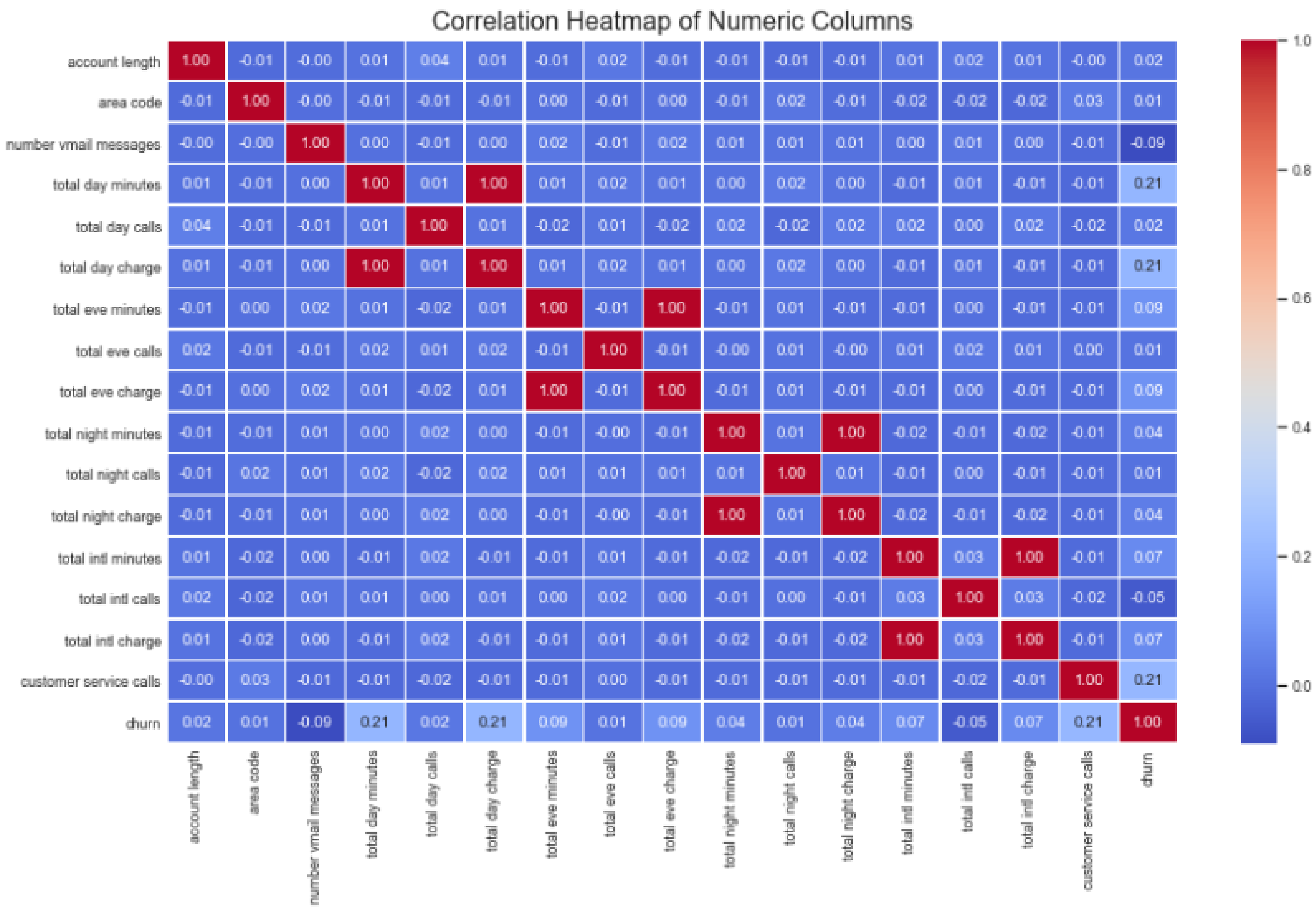
This included information about the customer's personal details, call habits and other churn indicators.

Data Understanding

The data was found to contain multiple variables of different types:

- account length int64
- area code int 64
- international plan object
- voice mail plan object
- number vmail messages int64
- total day minutes float64
- total day calls int64
- total day charge float64
- total eve minutes float64
- total eve calls int64
- total eve charge float64
- total night minutes float64
- total night calls int64
- total night charge float64
- total intl minutes float64
- total intl calls int64
- total intl charge float64
- customer service calls int64
- churn bool

Data Understanding



As you can see, there isn't a lot of perfect correlation among the variables except the few exceptions of minutes and charges.

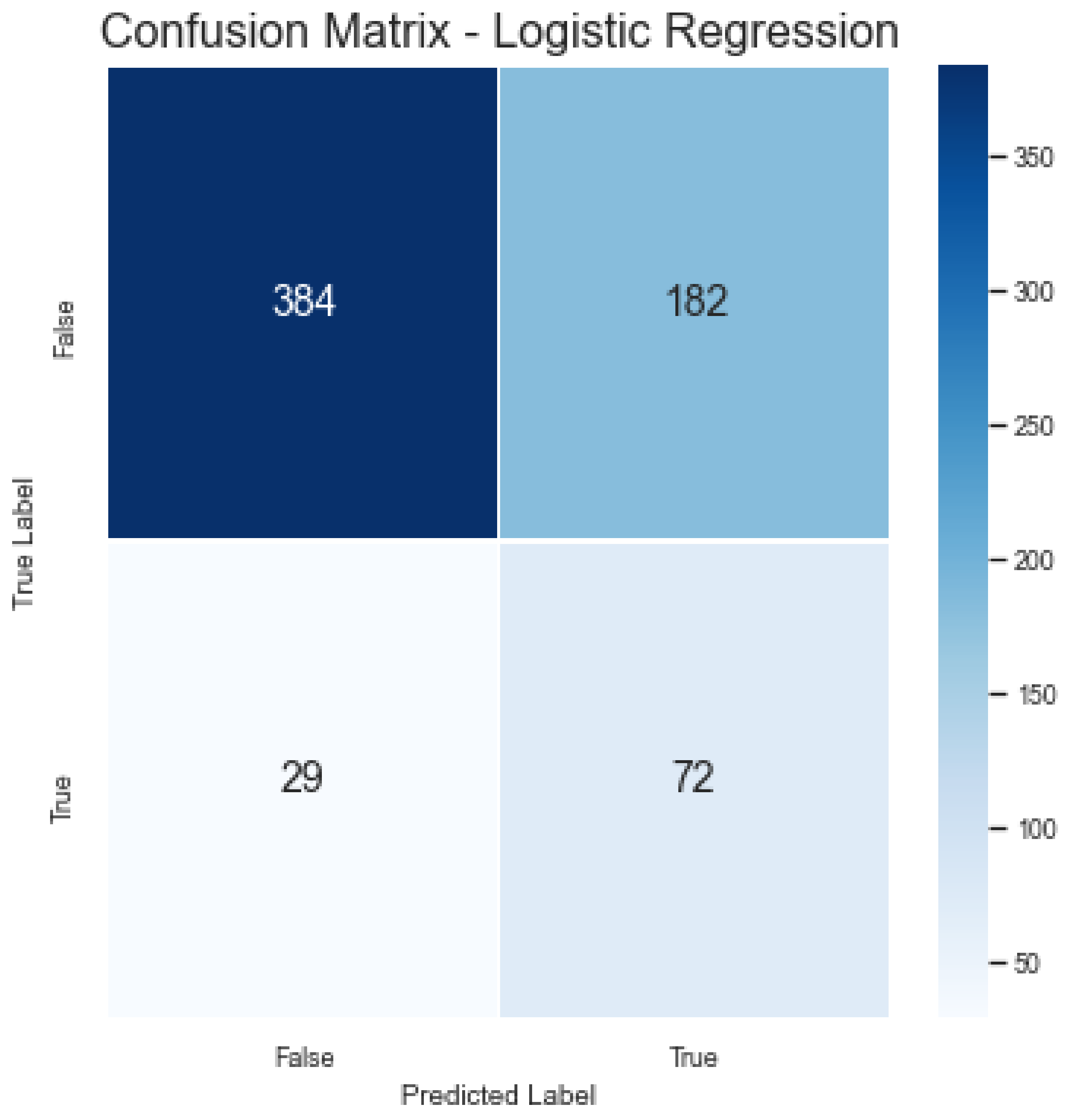
This however is to be expected as the charges are determined directly by the minutes spent on the call.

Objectives

The main objectives of conducting analysis were:

- To predict rate of customer churn
- To determine customer churn drivers

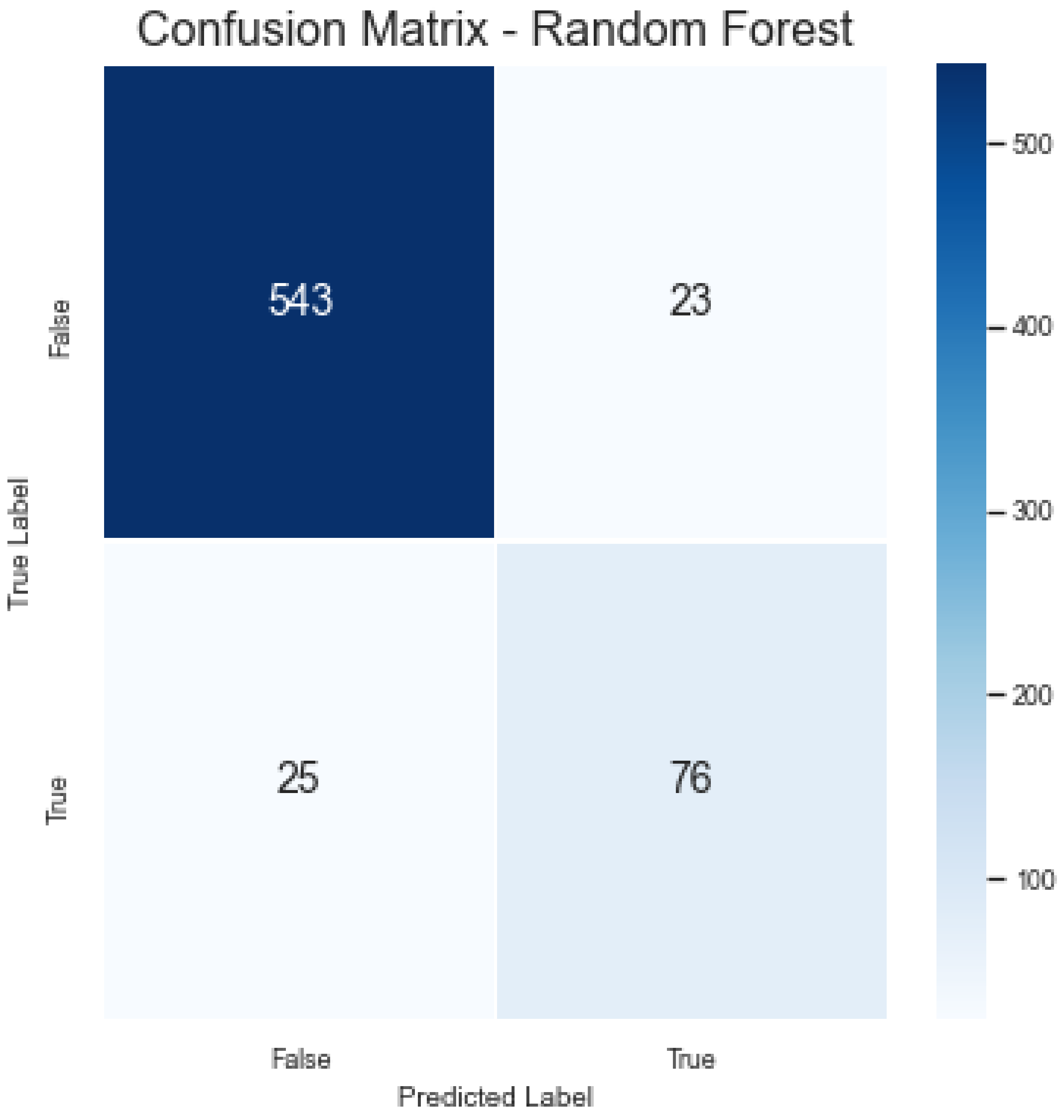
Logistic Regression Model



This model had a high number (182) of instances in which it incorrectly classified non-churn as churn while also correctly predicting 72 instances of churn.

The model is not yet optimal for use as it has a high degree of incorrect results needing further tuning.

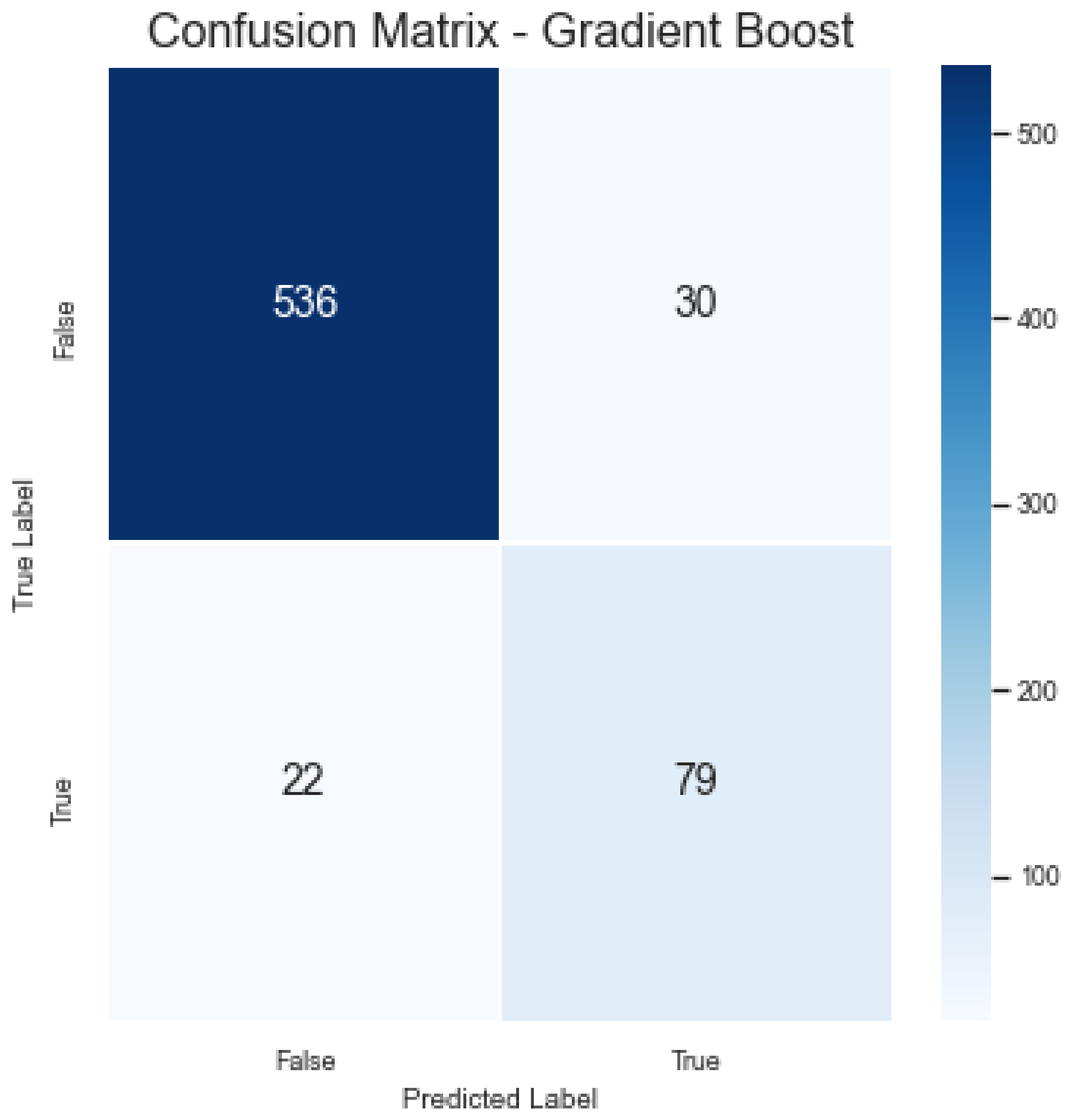
Random Forest Model



There were 76 correctly predicted instances of churn despite only 23 instances of non-churn incorrectly classified as churn.

This means this is a better model but the 25 false negatives indicate potential missed opportunities for getting customers at risk of churn.

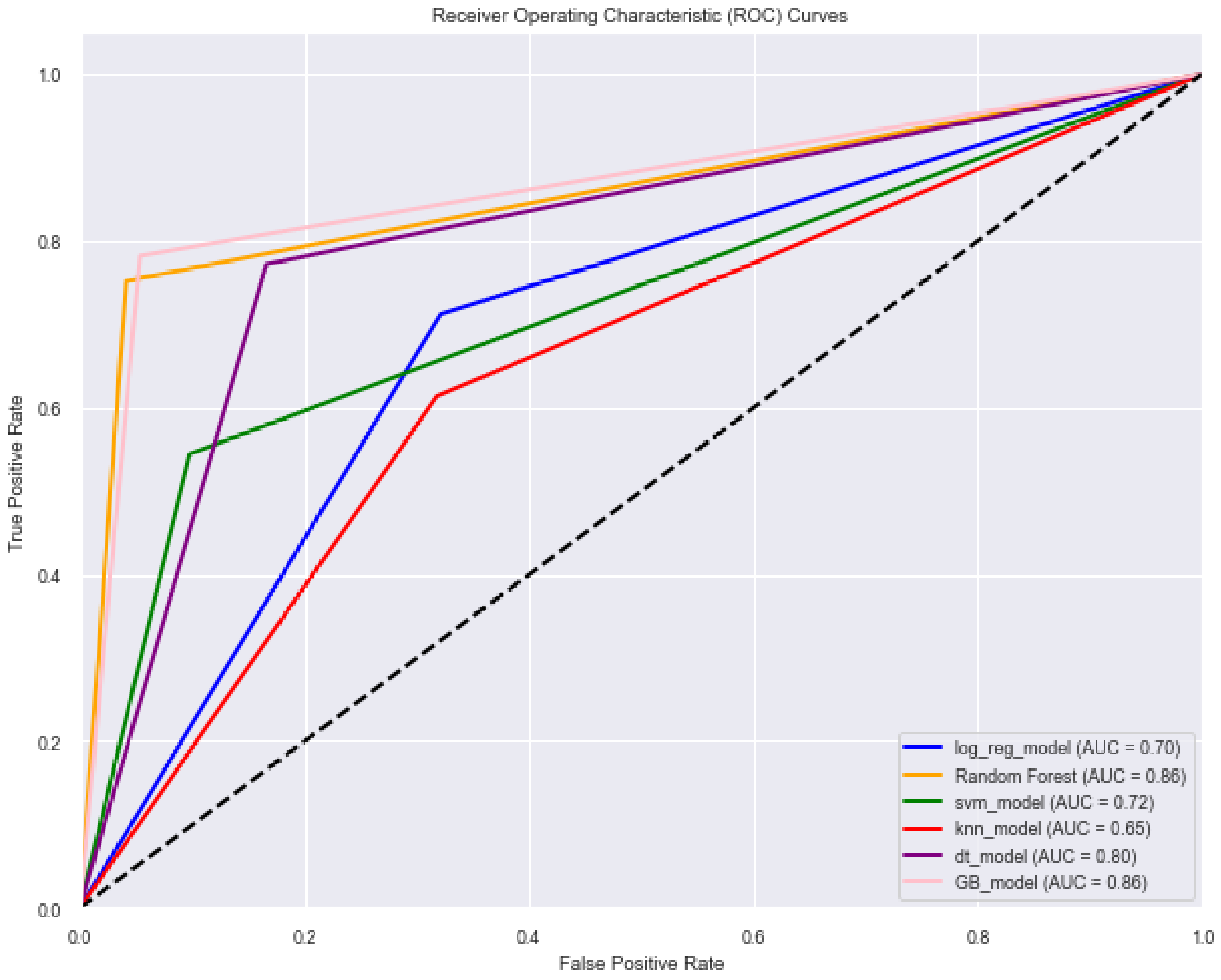
Gradient Boost Model



The Gradient Boost model predicted 79 instances of customer churn with only 30 instances of incorrect predictions.

This model produced the highest accuracies so was chosen for further fine tuning.

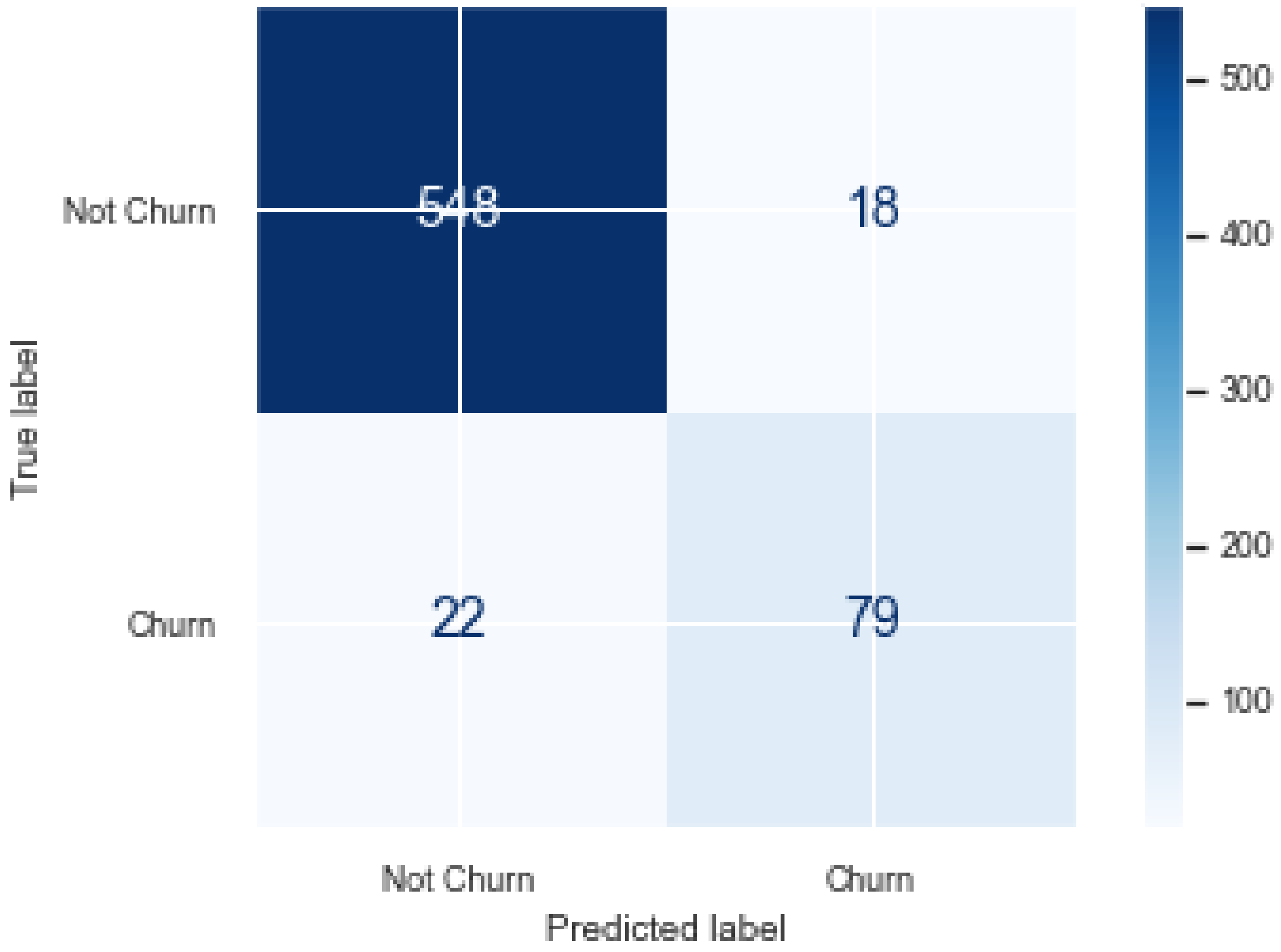
Evaluation(ROC)



As per the Receiver Operating Characteristic curve, the Random Forest and Gradient Boost model have the best predictive performances shown by their high AUCs.

The knn and logistic regression models have the poorest predictive abilities.

False Positive - False Negative Trade-off



In this trade-off, decrease of the false positives was prioritized to slightly increasing the false negatives.

This reduces the risk of losing customers incorrectly classified as non-churned aiding the business’goals.

Evaluation

Our model can correctly make predictions for approximately 93.7% of the customers, indicating that the model's predictions were accurate for the majority of the customers. Out of all the customers predicted as churned, approximately 69.44% of them actually churned, indicating that when the model identified a customer as churned, it was correct around 69.44% of the time.

Our model successfully captured about 86.96% of the customers who truly churned. The F1 score of 79.2% indicates that our model achieved a balanced trade-off between correctly identifying churned customers and minimizing false predictions.

Conclusion

In conclusion, both the Random Forest model and the Gradient Boost model show good performance in predicting customer churn. However, the Random Forest model has a slightly lower testing accuracy and precision compared to the other model. This suggests that the GB model may have a better balance between false positive and false negative predictions.

The trade-off between identifying as many churn cases as possible (high recall) and minimizing false positive predictions (high precision) is necessary for our analysis as it improves the predictive performance of our model and serves the objective of our stakeholder.

The features that contribute the most to whether a customer churns or not include 'customer service calls', 'total day charge', 'total day minutes', 'total international calls', and 'total eve minutes'.

Recommendations

- Assess the Pricing Structure: Syria Tel should consider examining charges and researching methods for optimizing pricing plans or introducing flexible pricing options to satisfy the different needs of customers.
- Targeted Retaining Strategies: According to the model's projections, the organization should focus on customers who are more likely to churn. Identifying these clients ahead of time allows the organization to interact with them proactively, offering specific rewards, discounts, or upgraded services to encourage continuous loyalty.
- Customer Sectioning: Syria Tel should separate its customers depending on their churn propensity, allowing it to better customize its marketing and retention efforts.

Next Steps

- Improving False Negative Predictions: While the recall score suggests that our model identified a considerable proportion of churned consumers, there is still space for improvement in reducing false negative predictions. It is critical to identify these consumers in order to adopt proactive retention efforts and reduce turnover.

- Continuous tracking and evaluation: Regular model review and iteration are required to improve model performance.

Thank you.