Customer Churning Prediction in Telcom Industry

Flow of Work

1. Introduction & Business Understanding
2. Objectives
3. Data Description
4. Model Building Process
5. Model Evaluation & Results
6. Factors Influencing Customer Churn
7. Conclusion & Recommendations

Business Understanding

Understanding telecom customer churn is crucial for revenue stability. By analyzing data to predict churners, companies can tailor retention strategies, reduce churn rates, and foster loyalty. This proactive approach minimizes revenue loss and informs product development, enhancing competitiveness in the telecom market

Objectives

- Identify factors contributing to customer churn.
- Develop a predictive model to accurately forecast churn.
- Implement targeted retention strategies to reduce churn rates.

Stakeholders

- -Senior Management
- -Marketing Team
- -Customer Service Reps
- -Investors
- -Communication Regulation Bodies
- -Sales Teams
- -Finance and Accounting Team
- -Product Development Team
- -Social Media Team

Key Objectives

- What factors contribute to customer churn within SyriaTel's customer base?
- Are there identifiable patterns or trends in customer behavior that precede churn?
- How accurately can we predict which customers are likely to churn in the near future?
- What proactive measures can SyriaTel take to retain at-risk customers and minimize churn rates?

Data Description

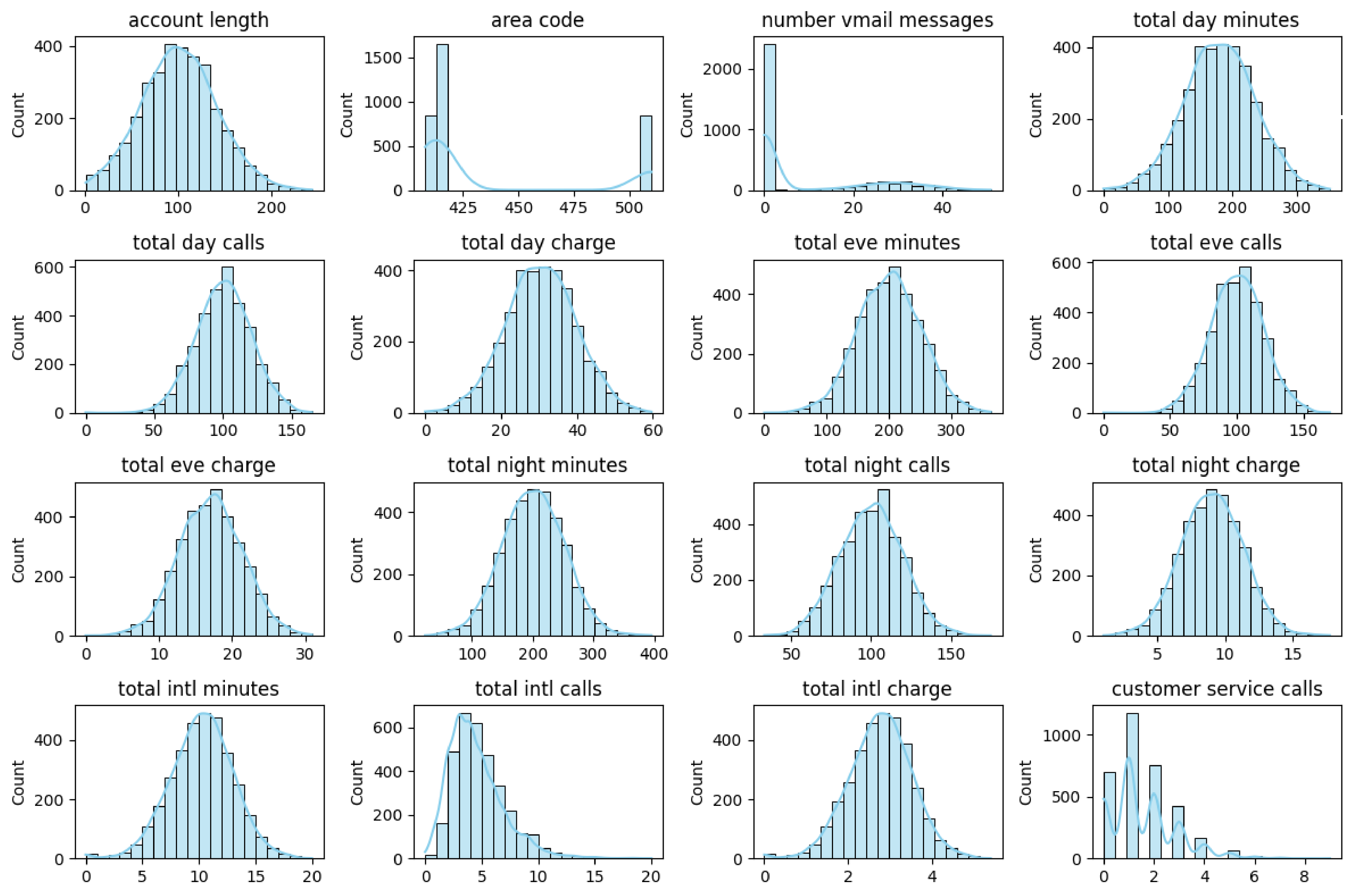
- State
- Account Length
- Area Code
- International Plan
- Voice Mail Plan
- Number of Voicemail Messages
- Total Day Minutes
- Total Day Calls
- Total Day Charge
- Total Evening Minutes
- Total Day Charge
- Total Evening Minutes
- Total Evening Calls
- Total Evening Charge
- Total Night Minutes
- Total Night Calls
- Total Night Charge
- Total International Minutes
- Total International Calls
- Total International Charge
- Customer Service Calls
- Churn

Libraries Used

Pandas
NumPy
Matplotlib
Seaborn
Plotly Express
StandardScaler
OneHotEncoder
Train_test_split
LabelEncoder
SMOTE
LogisticRegression

GradientBoostingClassifier
DecisionTreeClassifier
RandomForestClassifier
SVC
KNeighborsClassifier
Pipeline
Cross_val_score
Classification_report
Confusion_matrix
GridSearchCV
RandomizedSearchCV

EDA

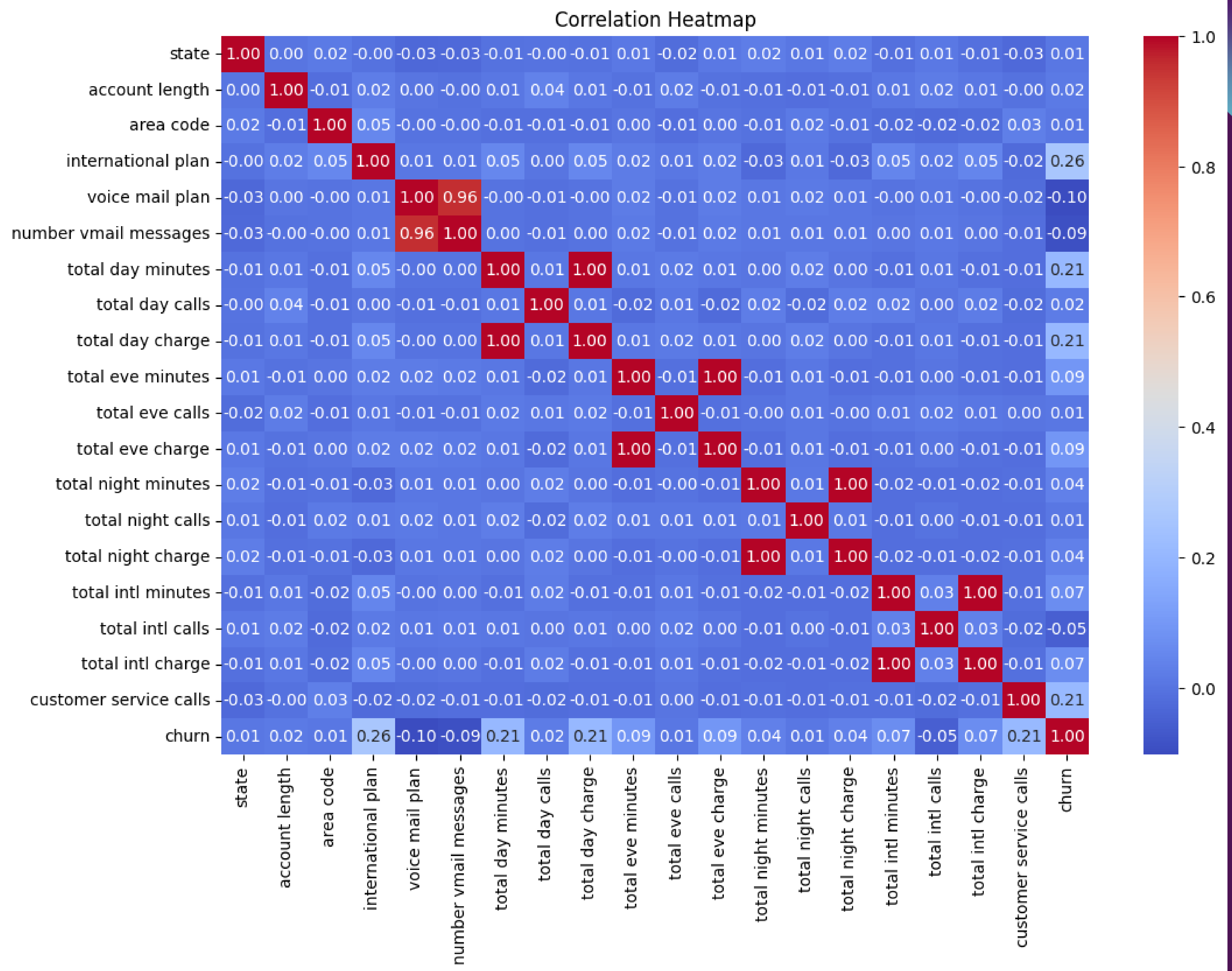


This is the distrbution of the numerical categories in the dataset

Data shows there is skewed distribution in Customer service calls

Total Minutes, Calls, Charge are nomally distributed

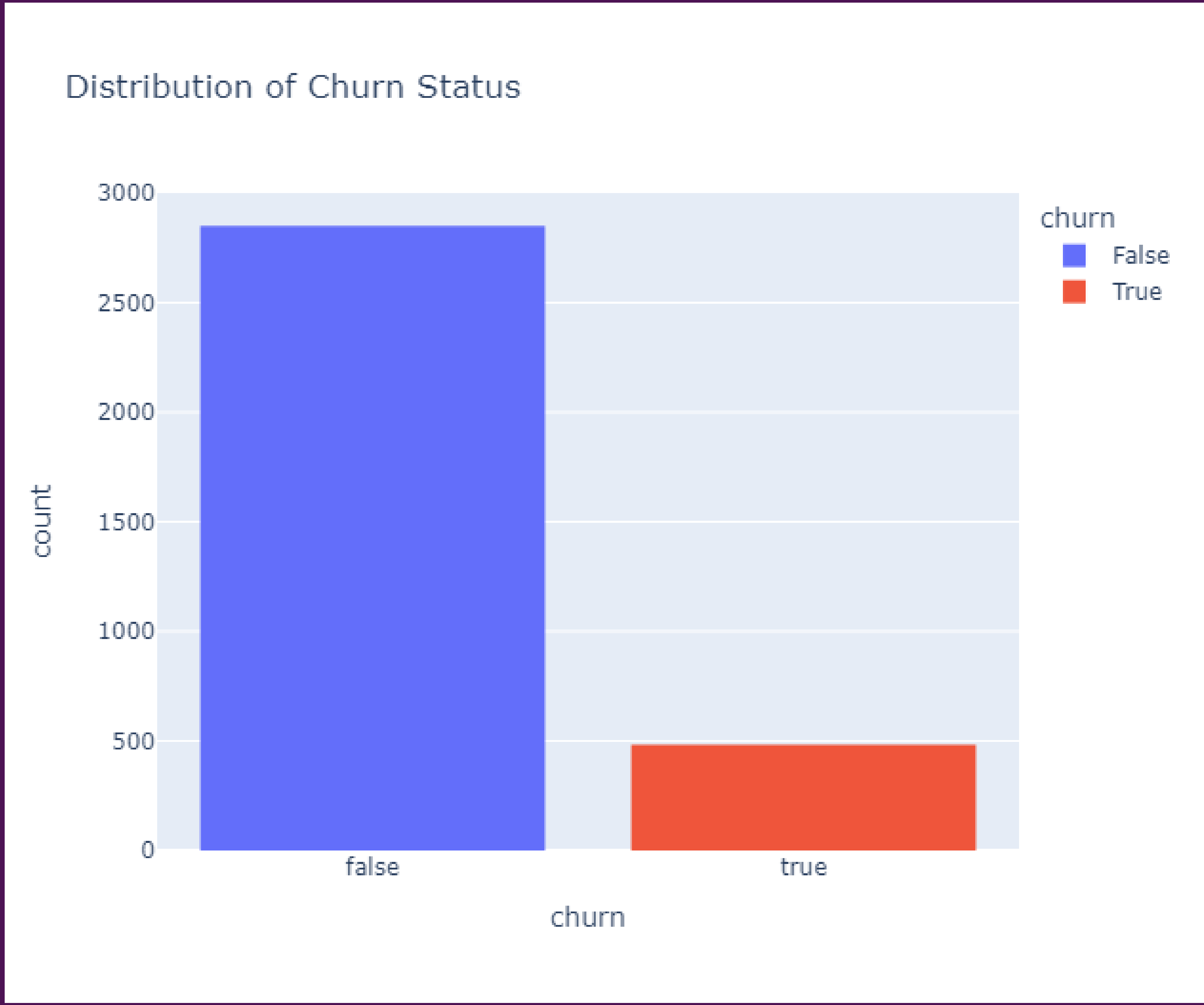
HeatMap Visuals



There is high correlation in international plan, customer services calls, total day minutes to churn.

Therefore these columns have an effect to the customer churn, and good to find out the influence on customer churn

Churn Status Distribution

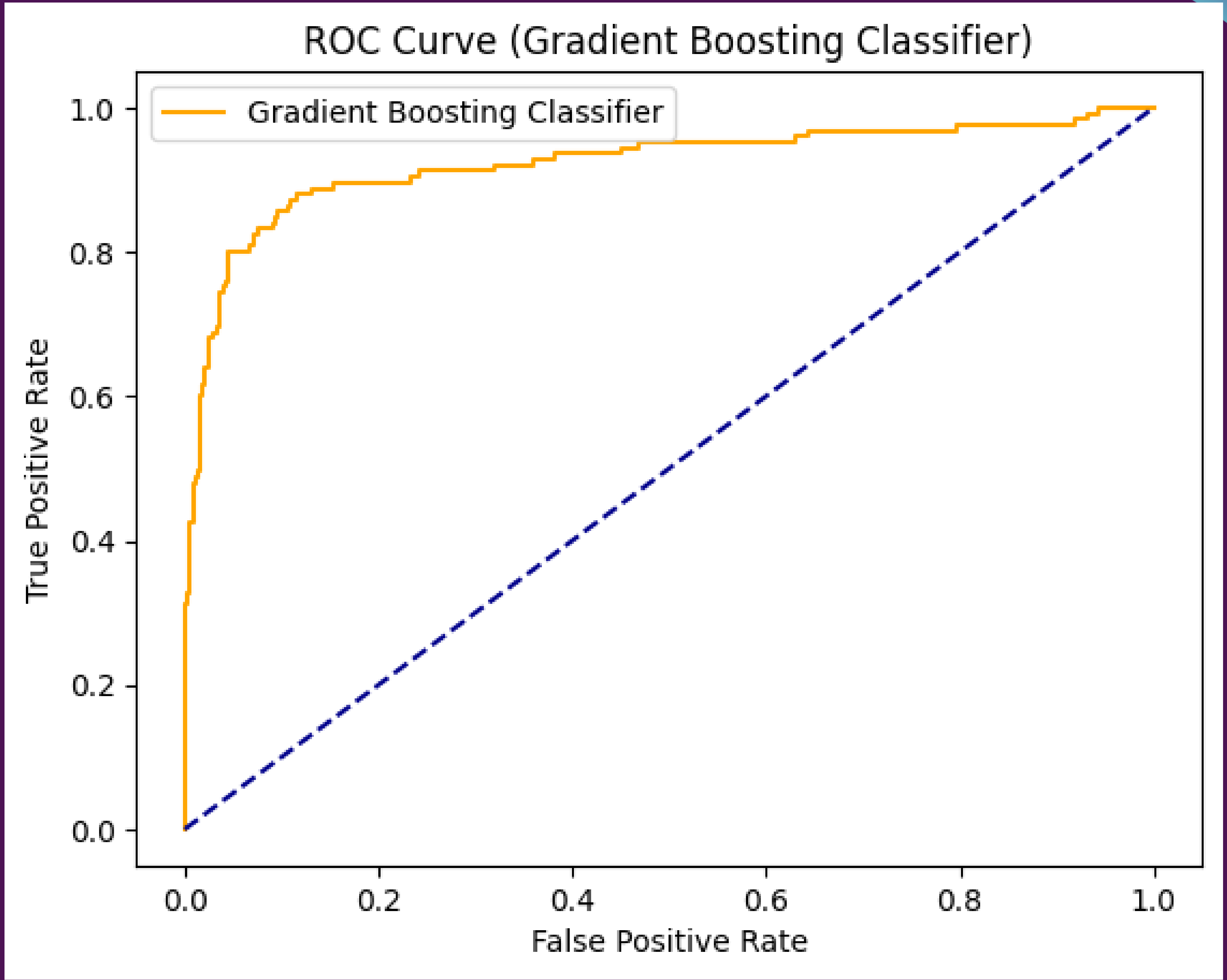


There is imbalance in the distribution of the customer churn.

The distribution of the data shows there is low customer churn in the dataset.

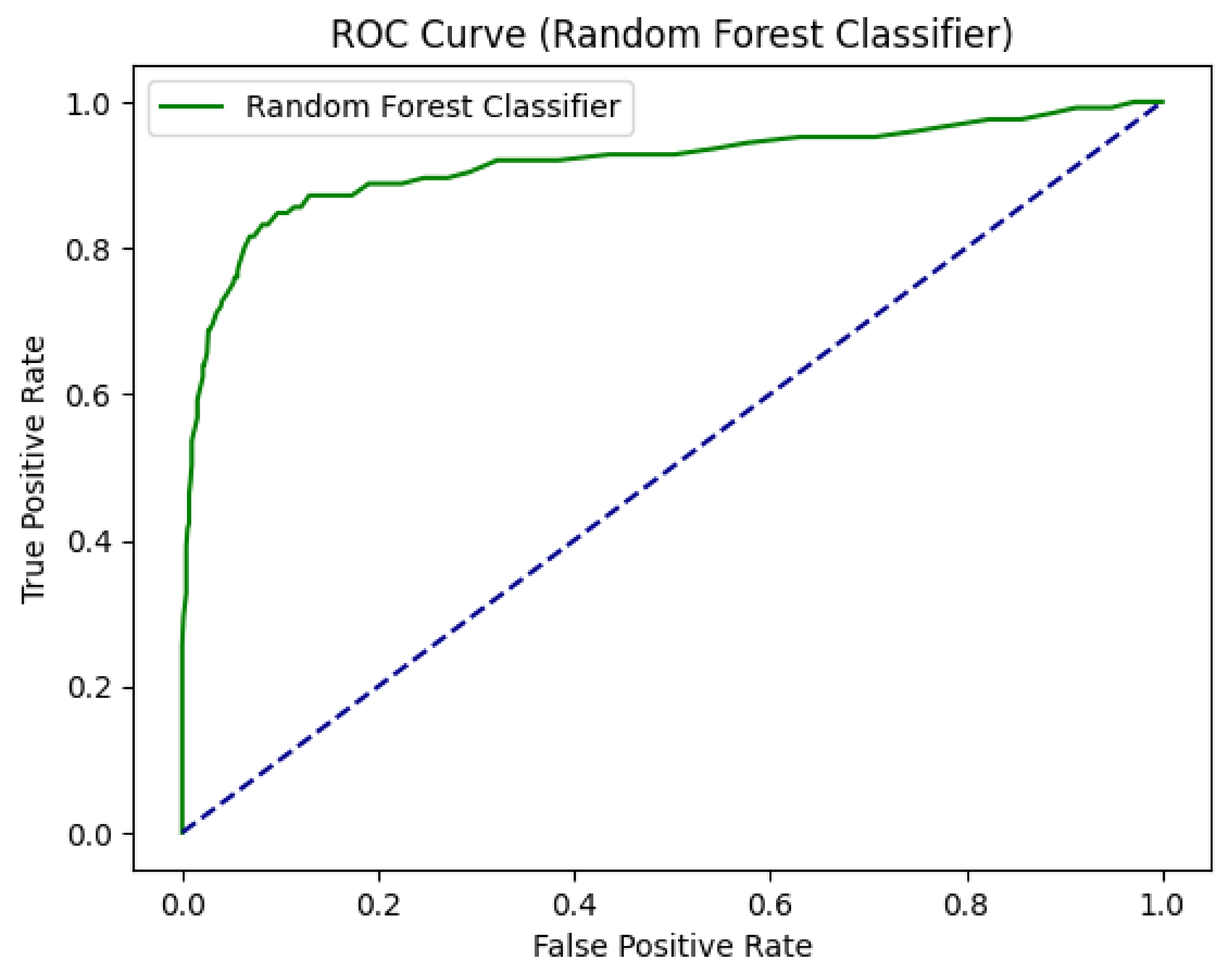
Reduce Imbalance in the churn status before modelling

Modelling



- Accuracy: 92.69%
- Precision: 78.07%
- Recall: 71.2%
- F1-score: 74.48%
- ROC-AUC Score: 91.46%
- The model exhibits strong performance in identifying churn, with a good balance between precision and recall.

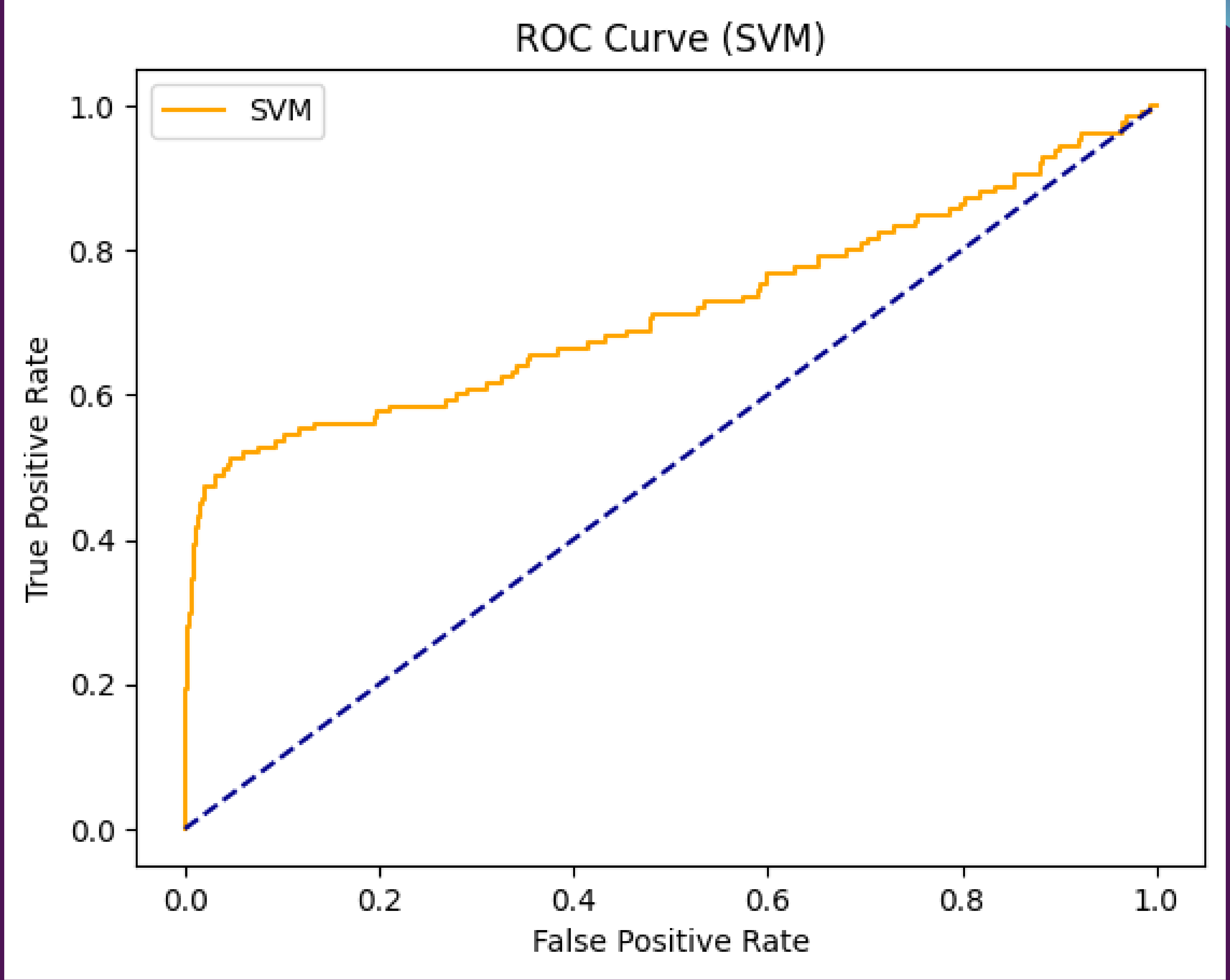
Modelling



- Accuracy: 93.17%
- Precision: 76.15%
- Recall: 79.2%
- F1-score: 77.65%
- ROC-AUC Score: 92.26%

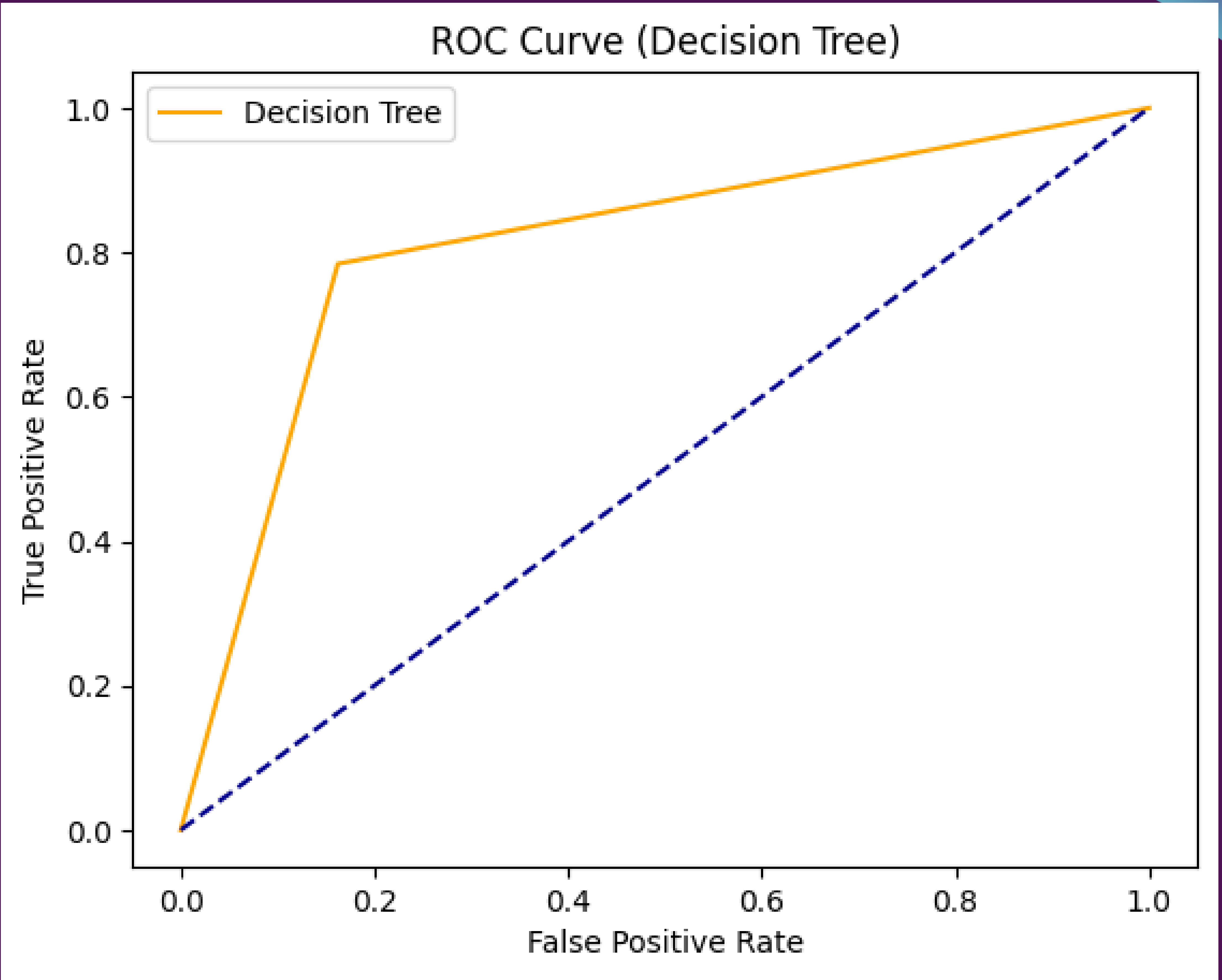
The model demonstrates strong predictive capabilities for identifying customer churn, with a balanced trade-off between precision and recall

Modelling



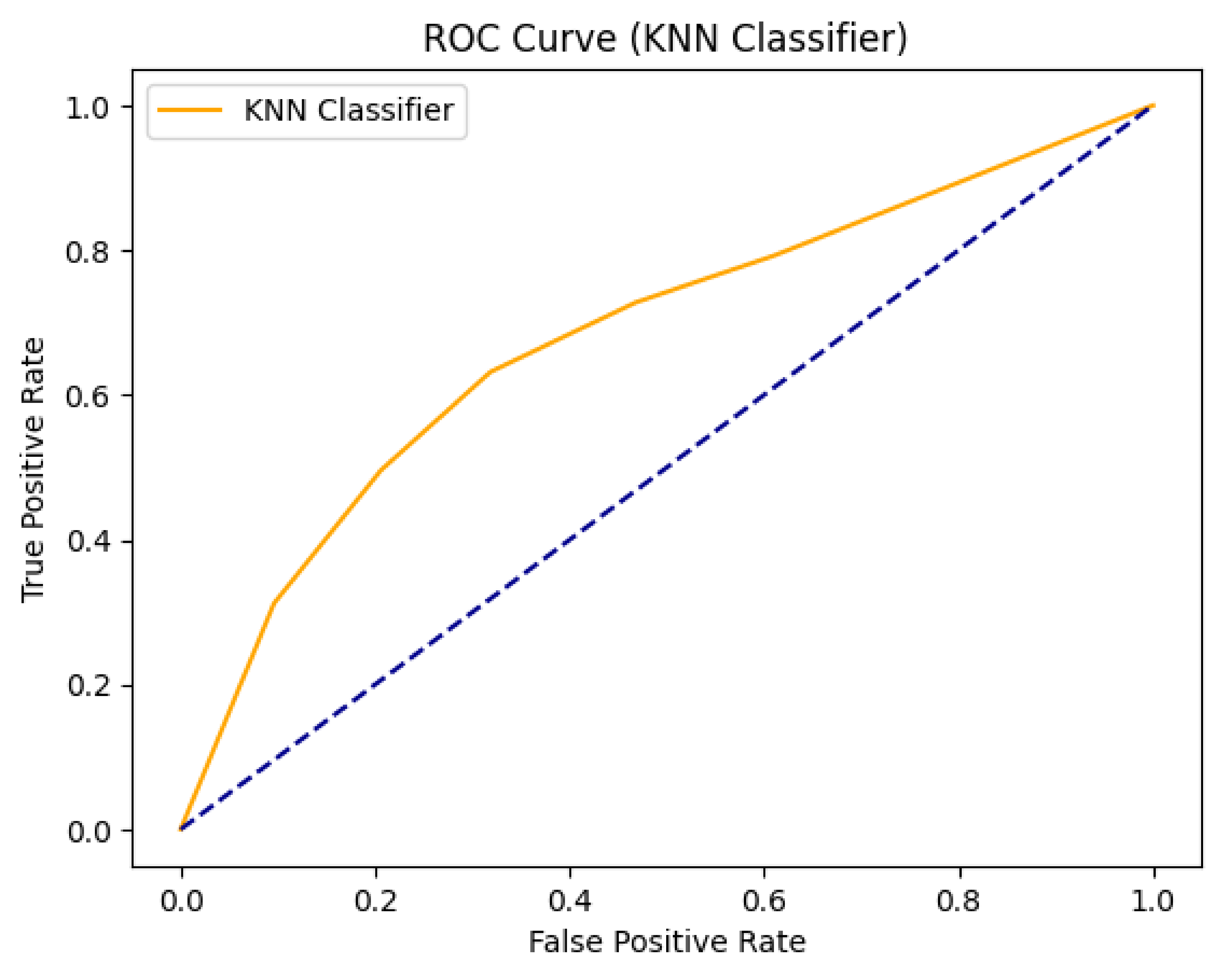
- Accuracy: 83.57%
- Precision: 45.95%
- Recall: 54.4%
- F1-score: 49.82%
- ROC-AUC Score: 71.54%
- The model's performance is moderate, with potential for improvement, particularly in precision and recall. There is room for enhancement to achieve better predictive accuracy.

Modelling



- Accuracy: 82.97%
- Precision: 46.01%
- Recall: 78.4%
- F1-score: 57.99%
- ROC-AUC Score: 81.09%
- The model exhibits moderate performance, with relatively high recall but lower precision

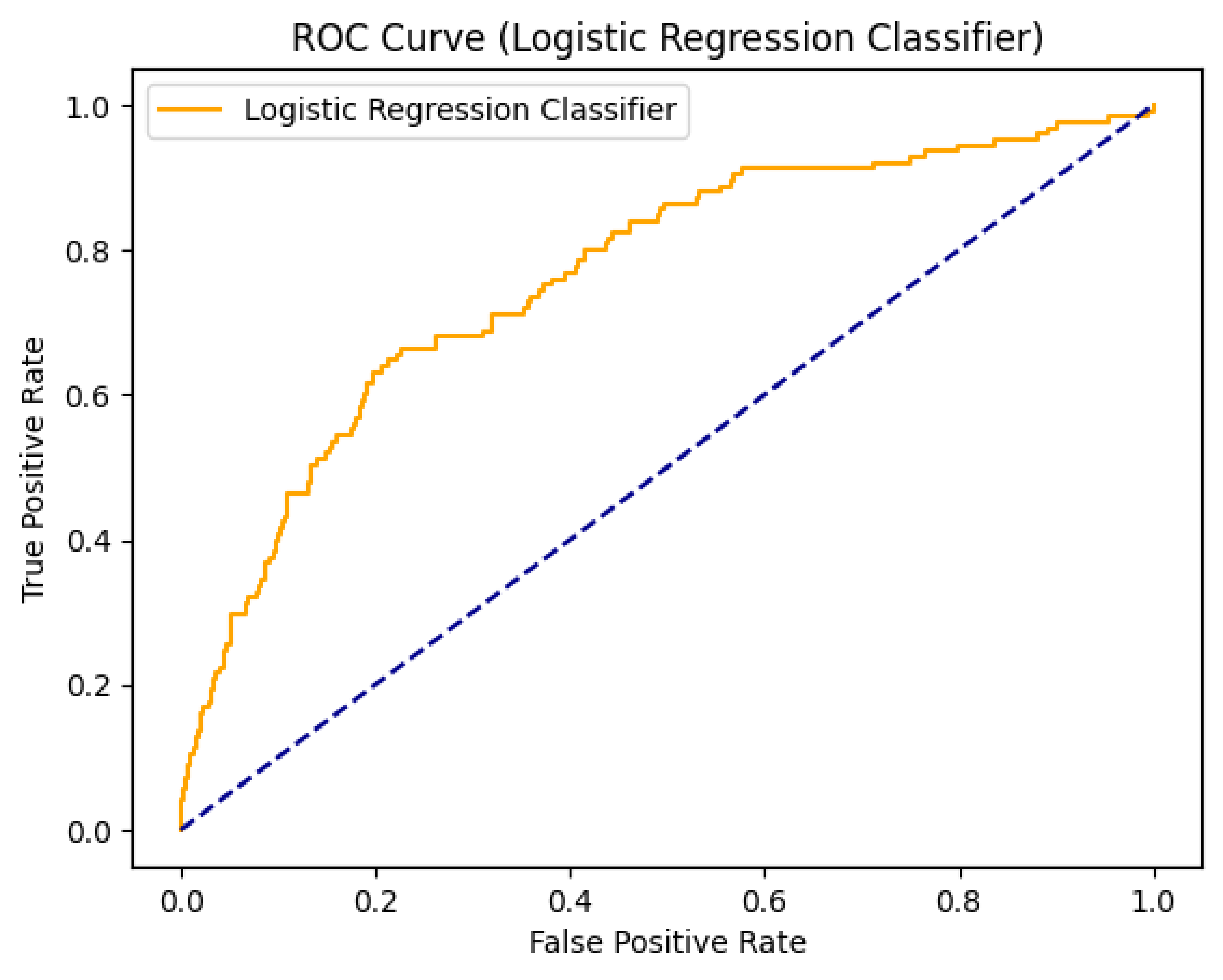
Modelling



- Accuracy: 67.39%
- Precision: 25.90%
- Recall: 63.2%
- F1-score: 36.74%
- ROC-AUC Score: 68.20%

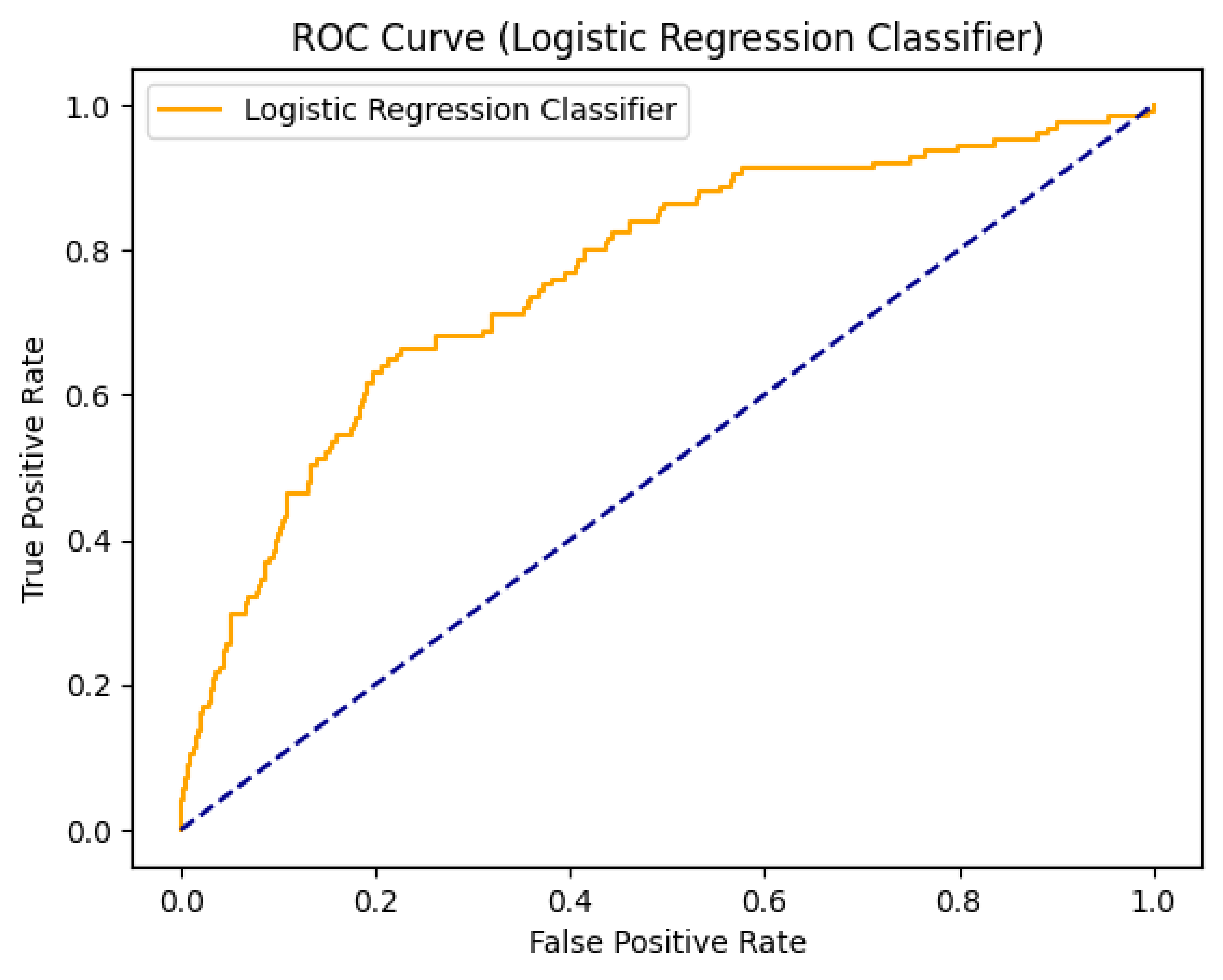
Model Performance is relatively low, with higher recall but lower precision. Improving precision would be essential to enhance the model's performance in accurately identifying churners while minimizing false positives

Modelling



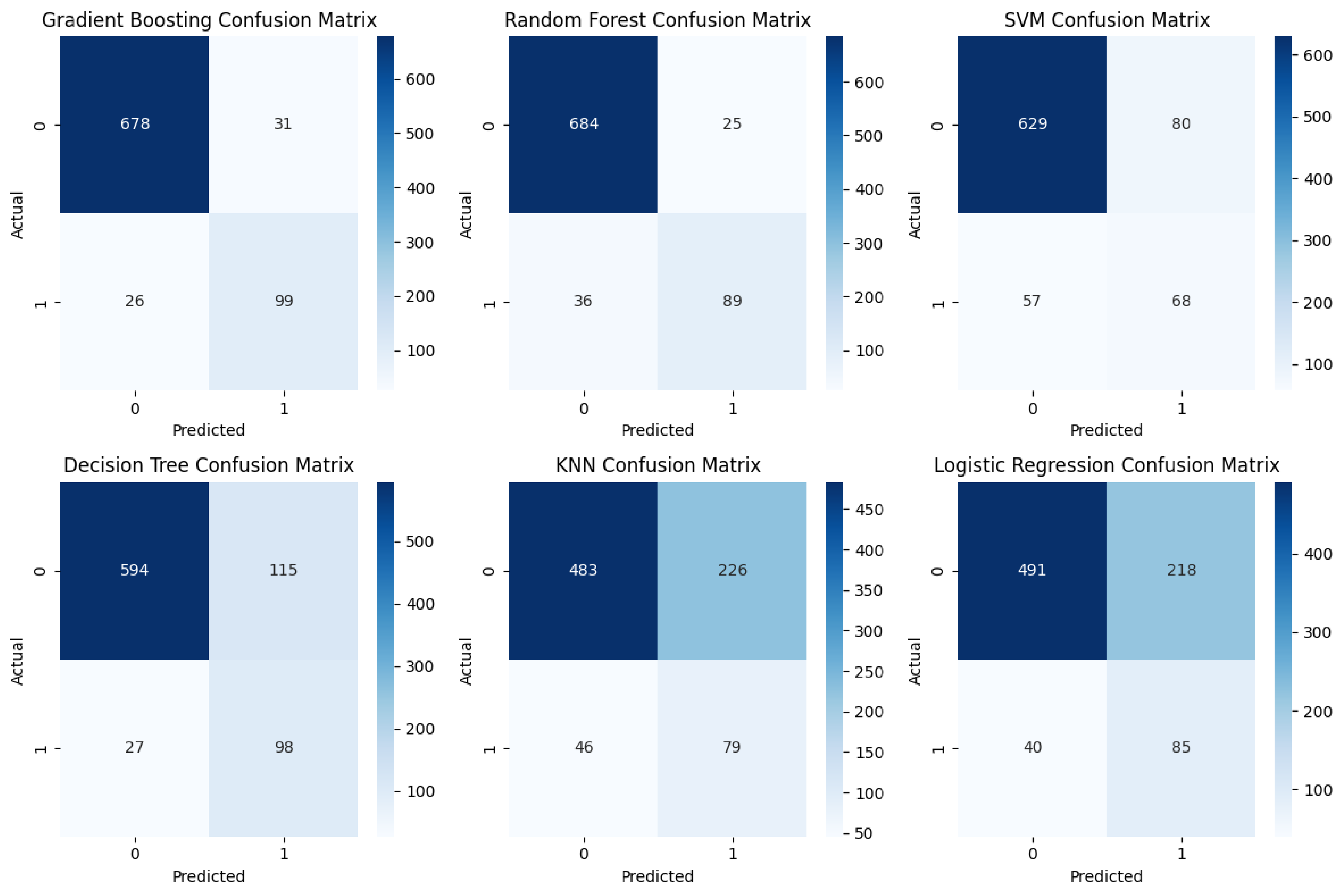
- Accuracy: 69.06%
- Precision: 28.05%
- Recall: 68.0%
- F1-score: 39.72%
- ROC-AUC Score: 76.22%
- The model shows moderate performance with higher recall than precision, suggesting it captures a good portion of actual churn cases but also has a significant number of false positive predictions.

Modelling



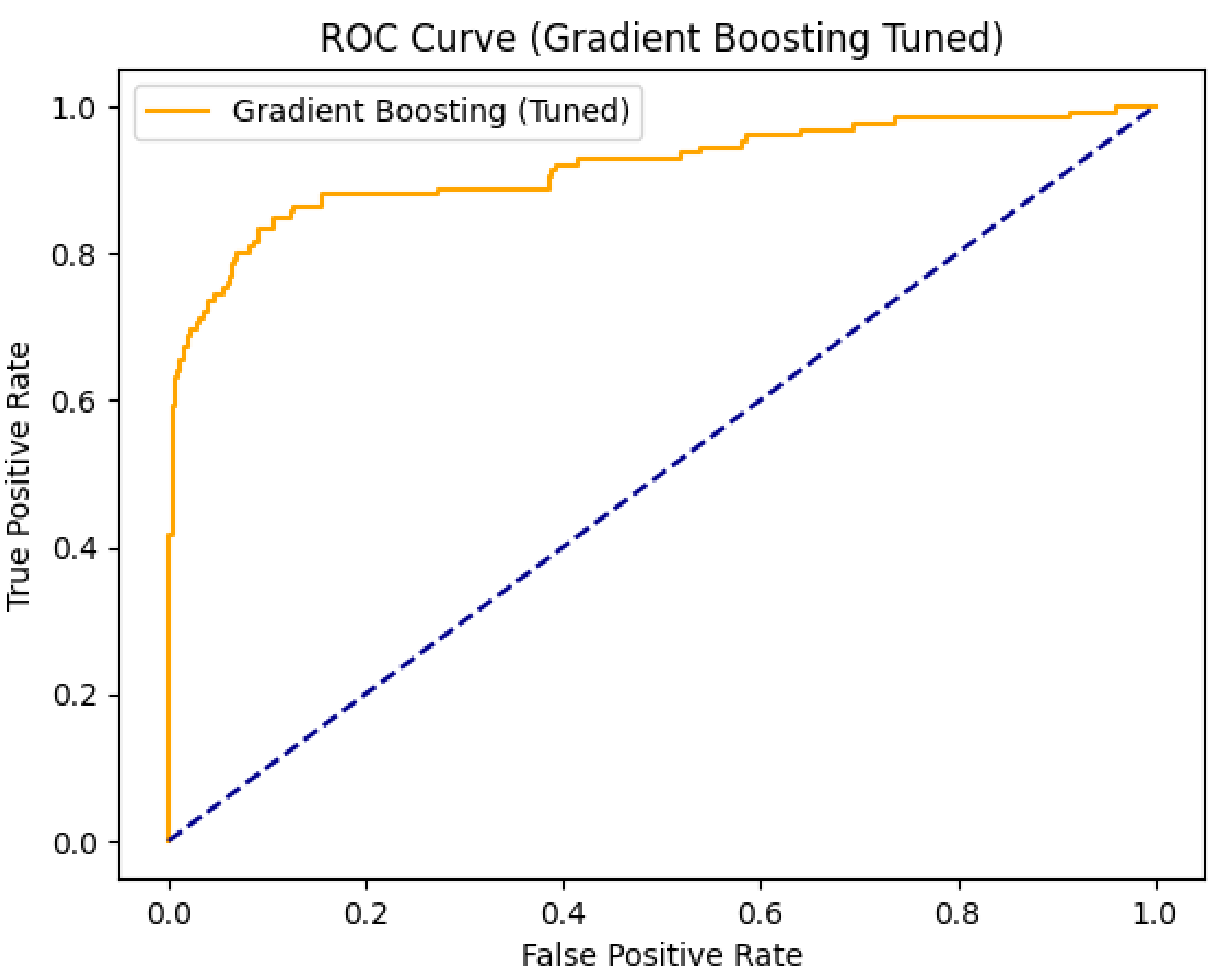
- Accuracy: 69.06%
- Precision: 28.05%
- Recall: 68.0%
- F1-score: 39.72%
- ROC-AUC Score: 76.22%
- The model shows moderate performance with higher recall than precision, suggesting it captures a good portion of actual churn cases but also has a significant number of false positive predictions.

Confusion Matrix



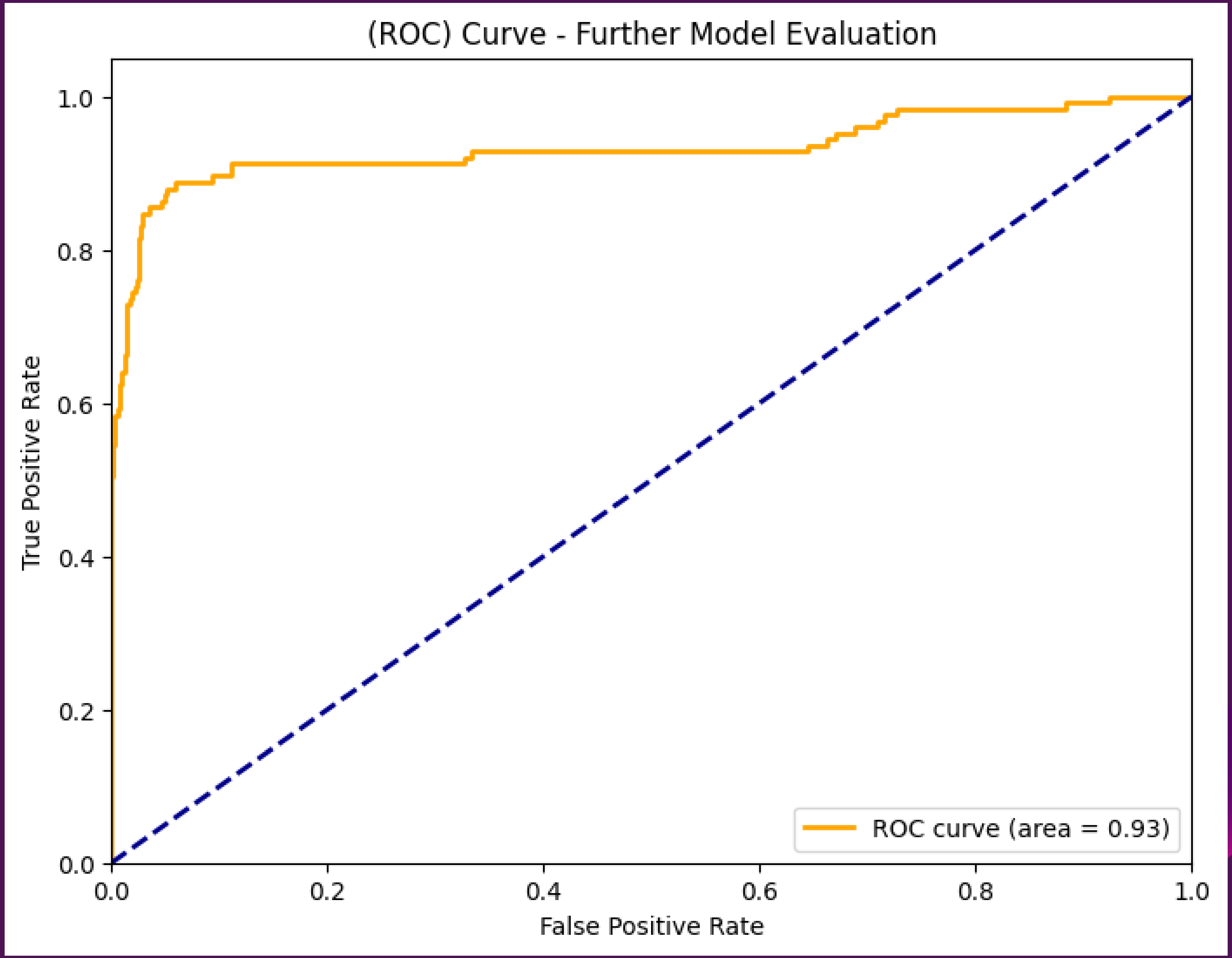
The Gradient Boosting and Random Forest classifiers exhibit the highest F1-scores, indicating superior performance. Their confusion matrices show a high number of correctly classified instances. Conversely, Logistic Regression and KNN classifiers have lower F1-scores and more misclassifications in their confusion matrices. Overall, the confusion matrix provides insights into the model's strengths and weaknesses, aiding in model evaluation and improvement.

Gradient Boosting Tuned



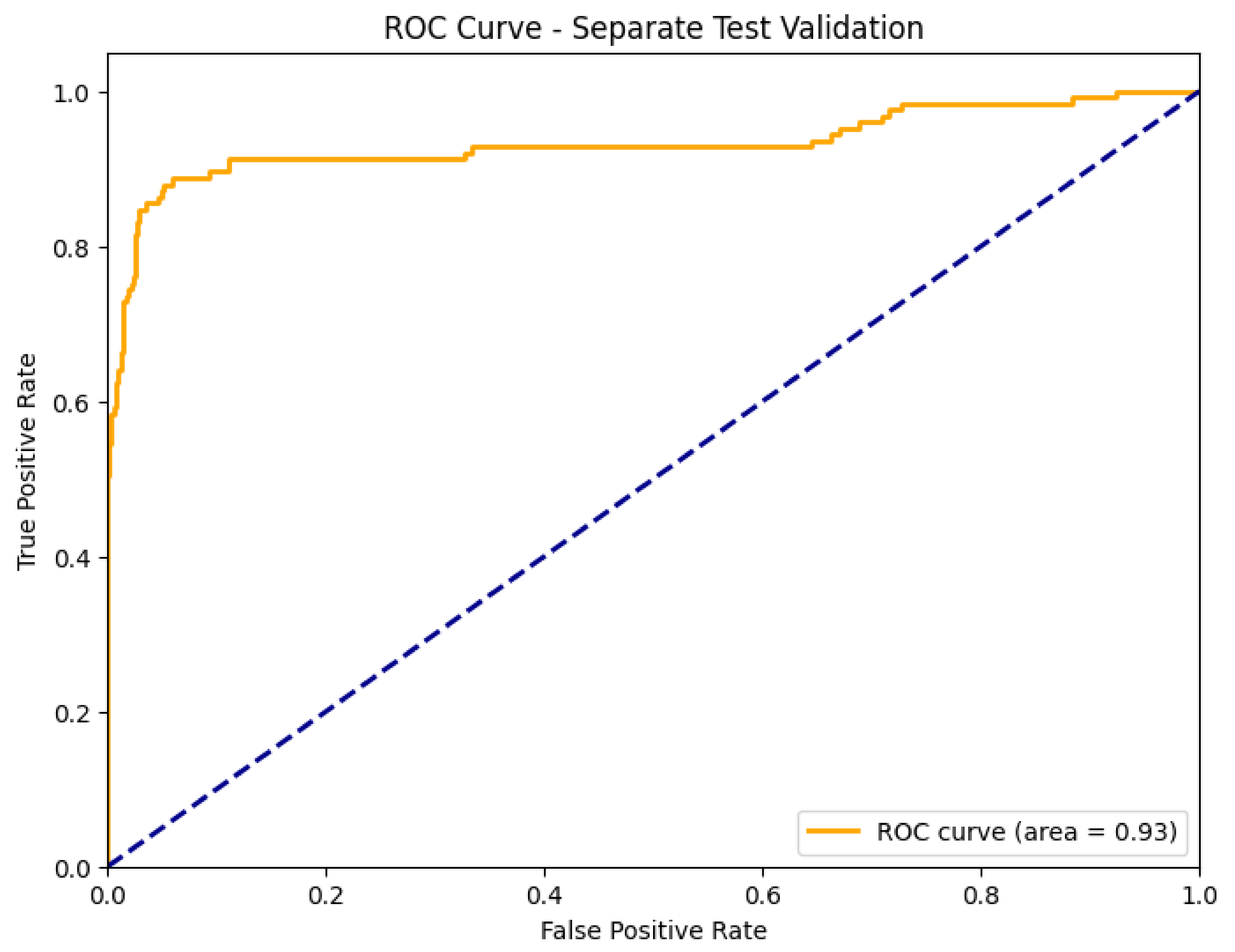
The model achieves an accuracy of 92.69% with a precision of 77.59% and recall of 72%. It shows a good balance between precision and recall, indicating strong predictive capabilities for identifying customer churn.

Further Evaluation



The model achieved an accuracy of 94.60% with a precision of 89.22% and recall of 72.8%. while maintaining a good balance between precision and recall. The F1-score of 80.18% indicates a harmonious blend of precision and recall, while the ROC-AUC score of 93.25% highlights its effectiveness in distinguishing between churn and non-churn instances.

Separate Test Validation



The model shows strong performance with 94.60% accuracy, 89.22% precision, 72.8% recall, and an F1-score of 80.18%. Its ROC-AUC score of 93.25% indicates effective churn prediction.

Factors Influencing Customer Churn

1. Service Quality: Address service disruptions and performance issues promptly to retain customers.

2. Competitive Pricing: Conduct market research and adjust pricing strategies to remain competitive.

3. Customer Support: Enhance responsiveness and effectiveness of customer support channels.

4. Billing and Payment: Streamline billing processes and offer flexible payment options to minimize errors and frustration.

5. Service Offerings: Innovate and expand service offerings to meet evolving customer needs and preferences.

6. Customer Experience: Invest in improving website, app, and store interactions to enhance overall customer satisfaction and loyalty.

Recommendations

1. Enhance Customer Engagement: Implement personalized communication and proactive support to strengthen customer relationships.
2. Target Retention Campaigns: Develop tailored offers and incentives for at-risk customers identified by the churn prediction model.
3. Improve Service Quality: Address common pain points and enhance service reliability to boost customer satisfaction.
4. Proactive Customer Service: Identify and address potential churn triggers early through proactive interventions.
5. Data-Driven Decision Making: Continuously monitor customer data to refine the churn prediction model and inform retention strategies.
6. Cross-Functional Collaboration: Foster collaboration across departments to align customer retention efforts.
7. Customer Feedback Mechanisms: Gather and analyze customer feedback to drive continuous improvement of the customer experience.