SYRIA

SYRIA TEL CUSTOMER CHURN

Outline

01 Project Overview

02 Business Understanding

03 Objectives

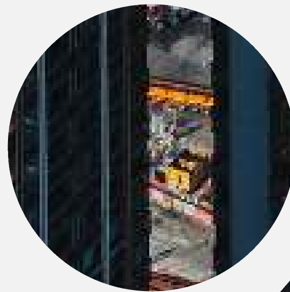
04 Data Understanding

05 Data Visualization

06 Data Modeling

07 Conclusions

08 Recommendations



Let's dive in

Project Overview

- SyriaTel faces customer churn, impacting revenue and market standing.
- To adapt, the company must understand attrition reasons, differentiate by offering superior services, and use predictive machine learning models.
- Build a classifier to predict whether a customer will ("soon") stop doing business with SyriaTel, a telecommunications company.

Business Understanding

SyriaTel, a telecommunications firm headquartered in Damascus, Syria, faces a significant challenge in managing customer churn, which poses a threat to its revenue and overall profitability. Customer churn refers to the situation where customers end their subscriptions, often switching to competitors or discontinuing the service altogether. Key factors contributing to churn include poor service experience, inadequate customer service, and the ease with which customers can switch providers. Addressing these issues is crucial for SyriaTel to retain customers and sustain its business. The primary aim of this project is to analyze patterns and reasons behind customer churn, enabling SyriaTel to take proactive measures to improve service quality, enhance customer support, and offer personalized solutions. By leveraging data-driven insights, SyriaTel can make informed decisions, tailor services, and allocate resources effectively to reduce churn and improve customer satisfaction, leading to financial benefits.

Objectives

1) Identifying common characteristics and behaviors associated with customers who have churned

2) Evaluate the importance of different features

3) Develop models and validate their predictive capabilities

Data Understanding

Data science encompasses the process of analyzing a dataset to uncover its patterns, content, and attributes. This particular dataset comprises 3333 entries and 21 features, each describing a customer with various characteristics. The 'churn' column serves as the target variable for predictive modeling.

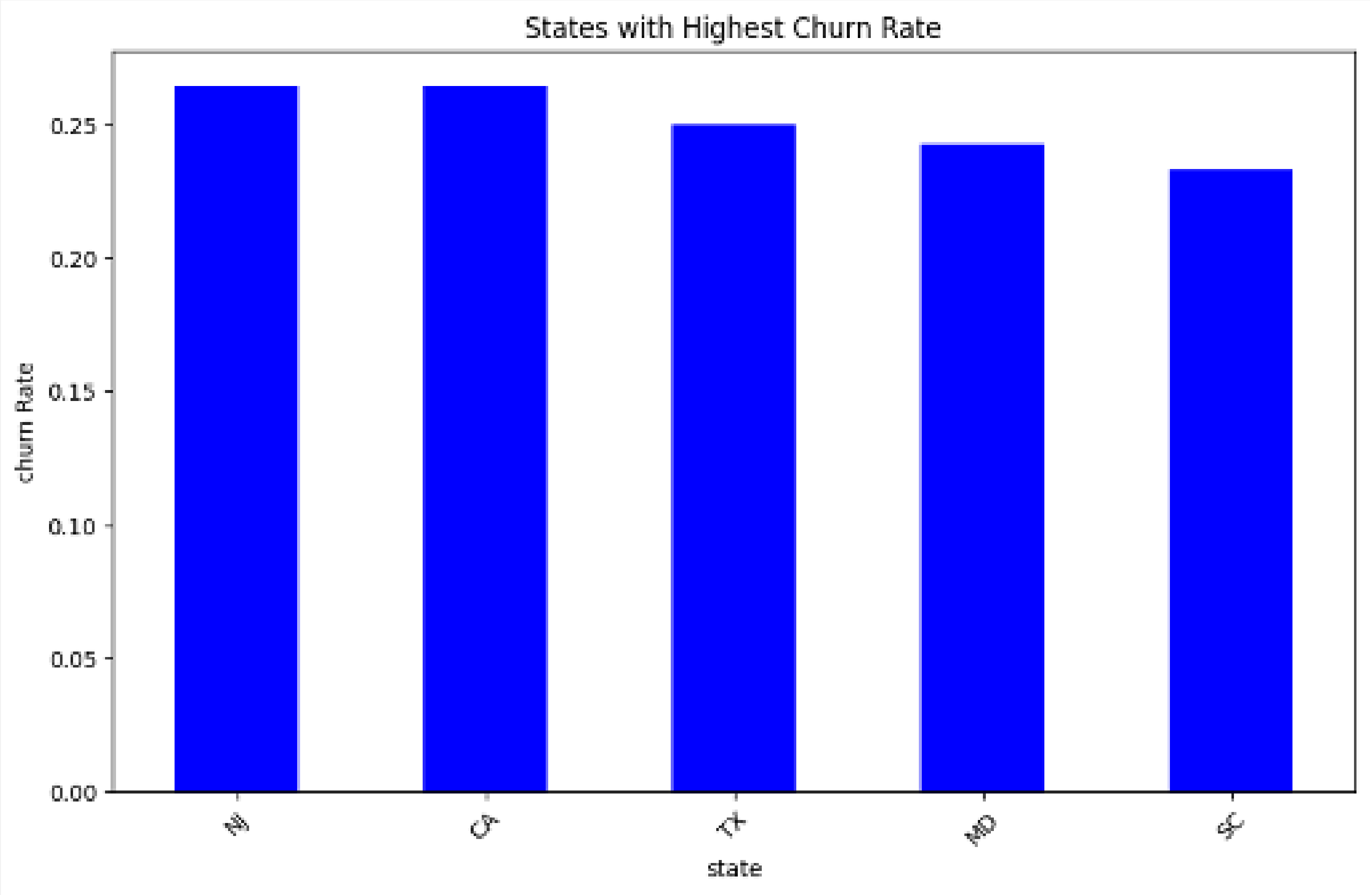
Data Understanding

The columns include;

state
account length
area code
phone number
international plan voice mail plan
number vmail messages
total day minutes, calls, charge
total eve minutes, calls, charge
total night minutes, calls, charge
total intl minutes, calls, charge
customer service calls
churn - our target

EXPLORATORY DATA ANALYSIS:

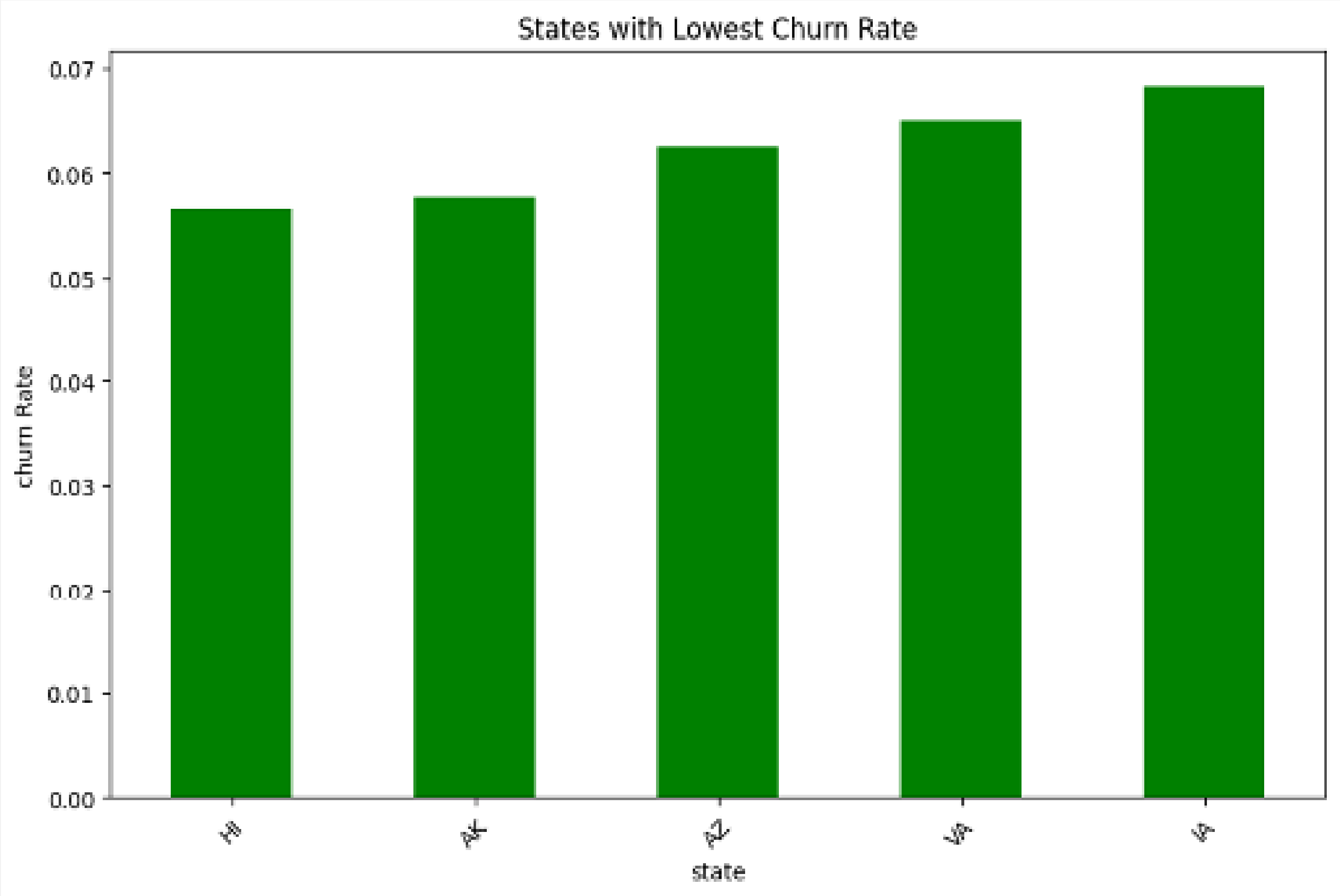
The top 5 states with the highest churn rate are:
1) NJ: New Jersey
2) CA: California
3) TX: Texas
4) MD: Maryland
5)SC: South Carolina



EXPLORATORY DATA ANALYSIS:

The top 5 states with the lowest churn rate are:

1) HI: Hawaii
2) AK: Alaska
3) AZ: Arizona
4) VA: Virginia
5) LA: Louisiana



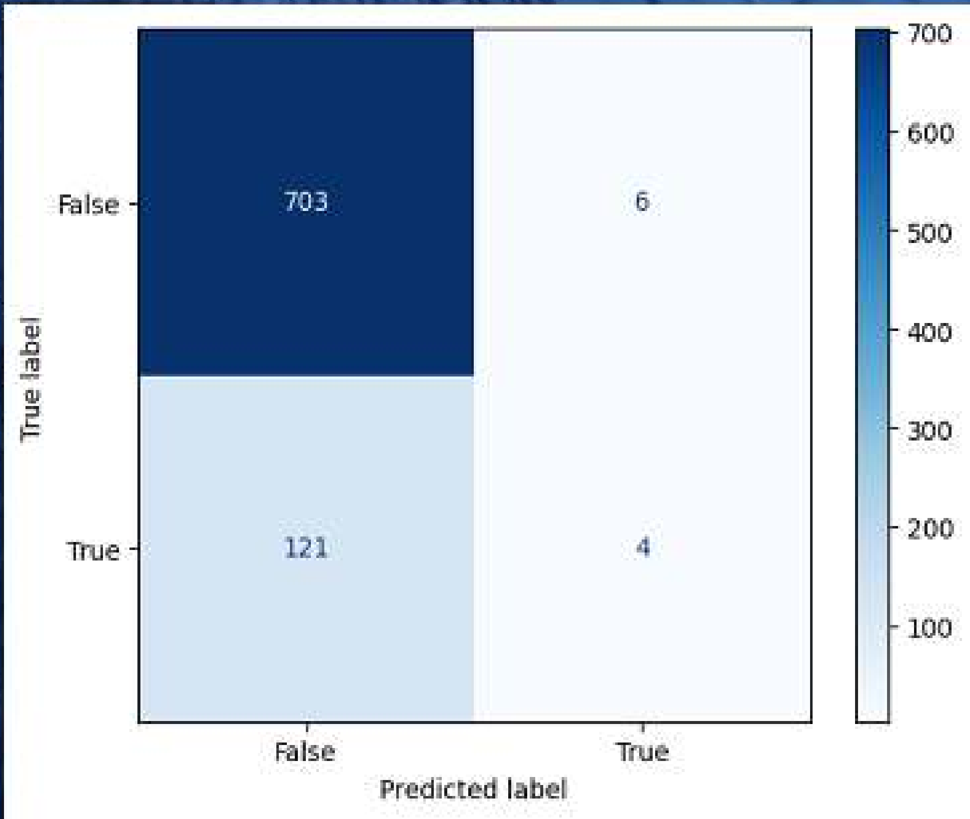
Data Modelling

RESULTS AND EVALUATION

Among the various models tested, including Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, Random Forest Classifier, and XGBoost, the XGBoost model emerged as the top performer overall. It achieved the highest accuracy, with a test accuracy of 91.97% and training accuracy of 95.64%. Additionally, based on the Test ROC and AUC score, which assesses the model's ability to differentiate between positive and negative outcomes, XGBoost achieved a score of 90%, further confirming its effectiveness.

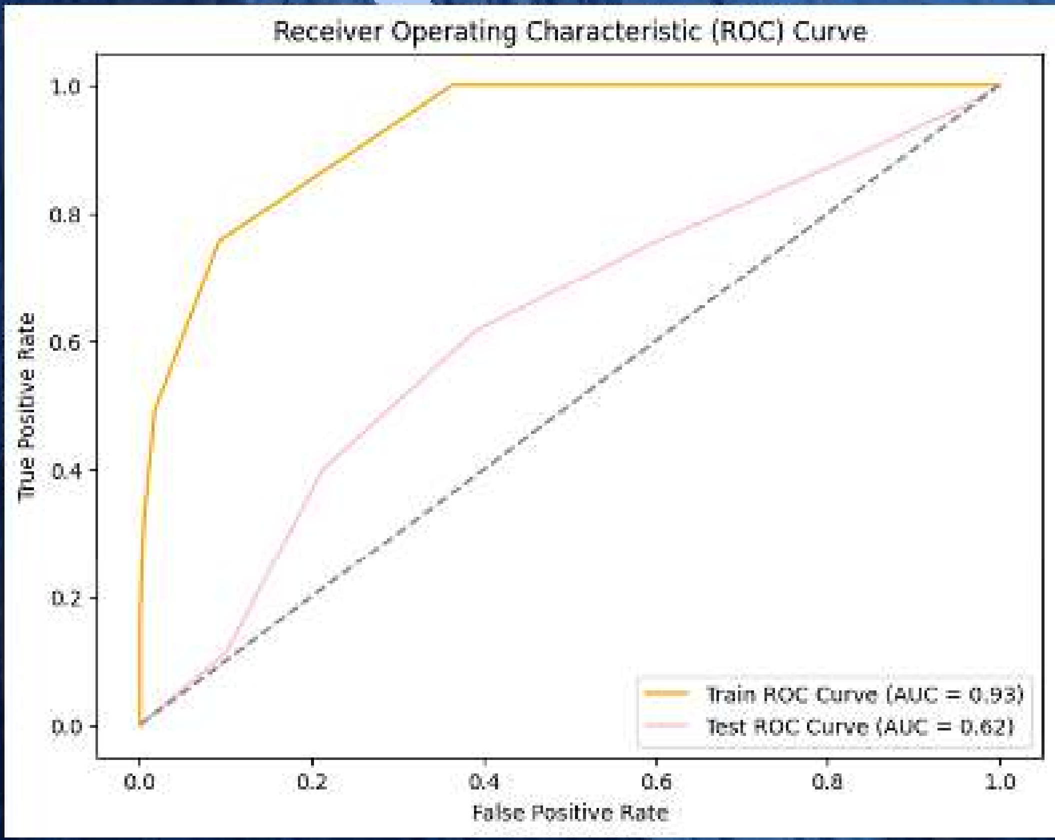
Logistic Regression

With training accuracy at 85.4% and testing accuracy at 84.8%, the model shows consistent performance, suggesting good generalization. However, its ability to predict churn is weak, with precision around 40% and recall approximately 4.2% for training and 3.2% for testing. Consequently, the low F1-scores indicate an imbalance between precision and recall. This baseline model's weaknesses can be addressed by evaluating other models for improvement.



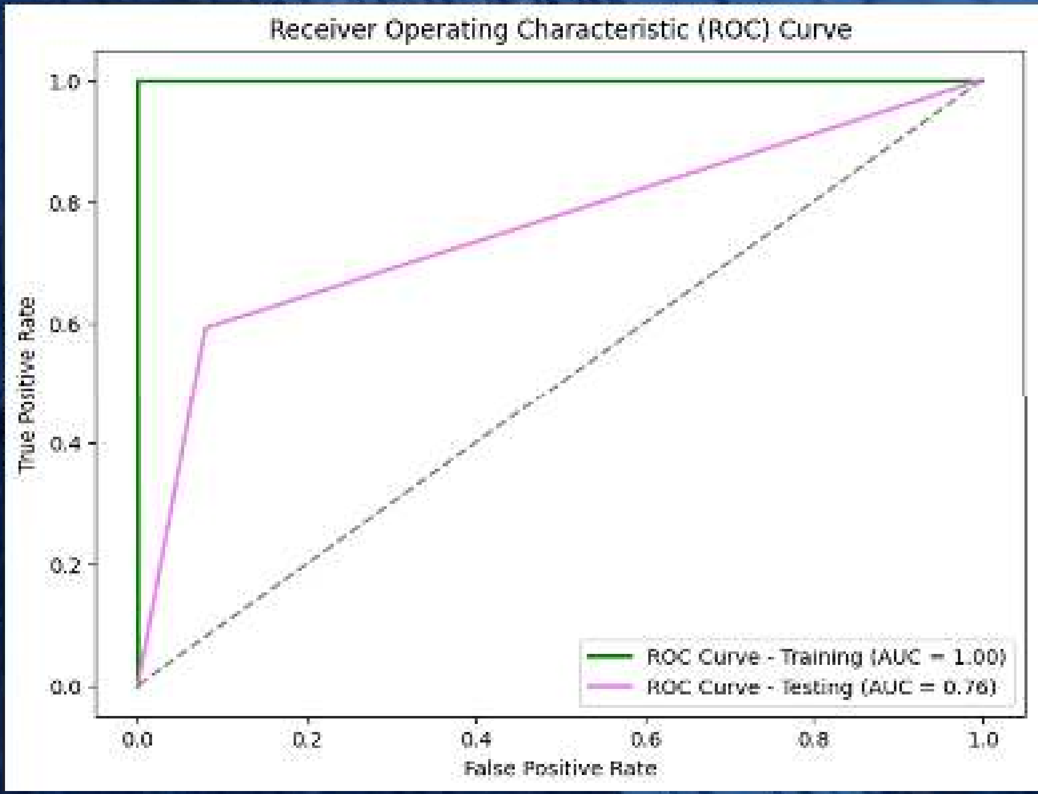
K – Nearest Neighbors

After optimizing the k-value, the model improves precision on the training set to 92.6%, with an accuracy of 89.8%. However, its recall remains relatively low at 31.3%. Though there's a slight enhancement in performance on the testing set, particularly in recall, precision remains low. Overall, while parameter tuning brings some improvements, the model still struggles to generalize to unseen data. Additionally, the AUC score drops notably on the testing set, indicating reduced discriminative performance compared to the training set.



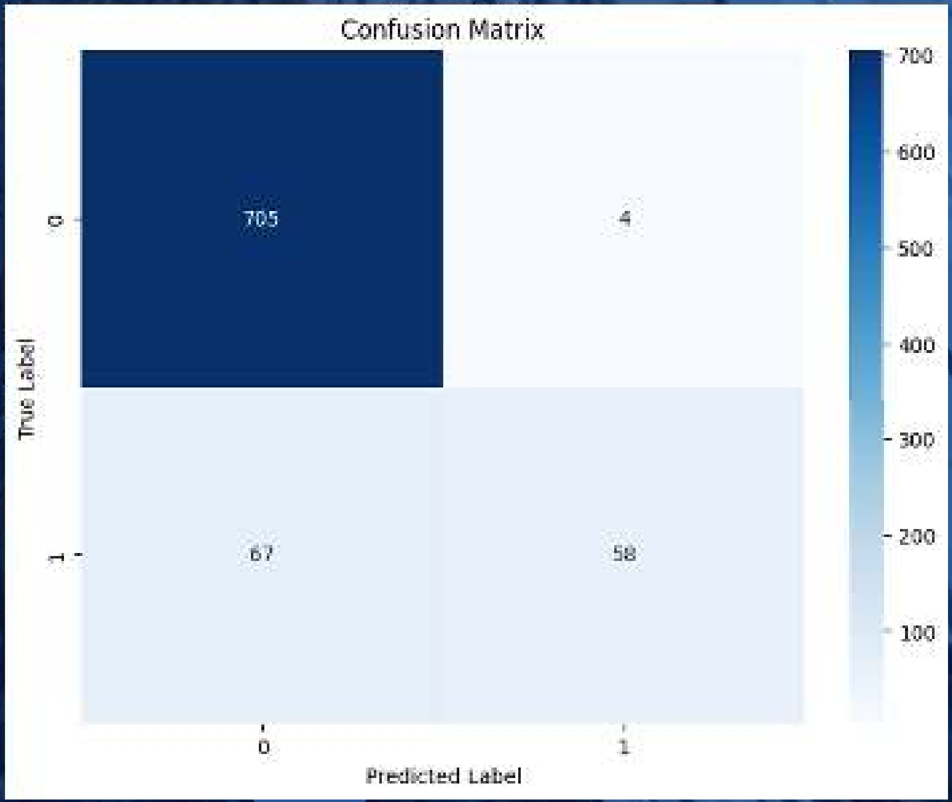
DECISION TREE

The model achieves perfect scores across all metrics on the training set, demonstrating accurate churn prediction within that data. However, on the testing set, there's a slight drop in precision and recall, although the accuracy remains high at 87.1%. The F1 score indicates a balanced trade-off between precision and recall, while the ROC AUC score suggests effective discrimination between churn and non-churn instances. This performance gap suggests potential overfitting, prompting the implementation of hyperparameter tuning to enhance generalization.



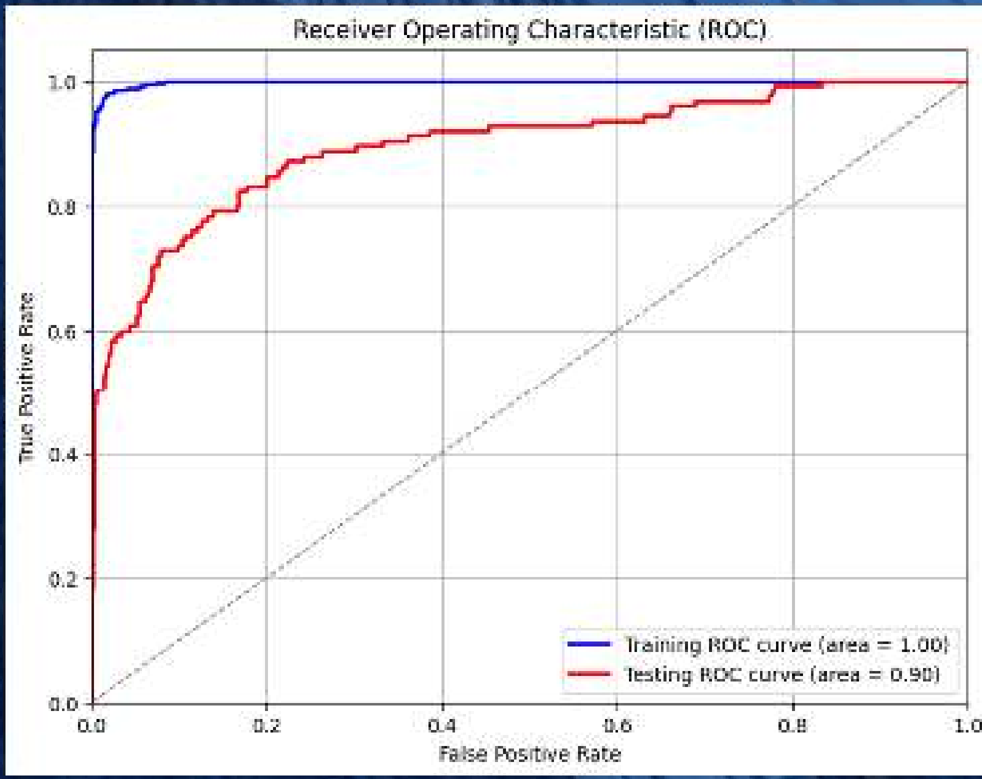
RANDOM FOREST CLASSIFIER

The model seems to be performing well in predicting True Negative and True Positive but higher number of False Negative.



EXTREME GRADIENT BOOSTING XGBOOST

Following tuning, the model shows improved performance on both training and testing sets. On the training set, precision remains high at 0.9960, but recall drops to 0.6983. However, accuracy increases to 0.9564, and the F1 score improves to 0.8210, indicating a better precision-recall balance. On the testing set, precision improves to 0.9143, though recall remains at 0.5120. The accuracy is high at 0.9197, and the F1 score increases to 0.6564, showing a better precision-recall balance. Overall, tuning enhances the model's metrics on both sets.



Conclusion

The main contributors to customer churn are:

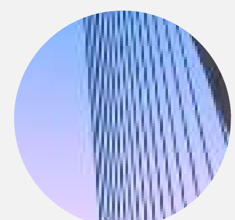
- Having an international plan, which increases the likelihood of churn.
- High usage of daytime minutes, associated with higher churn.
- Frequent customer service calls, indicating a higher chance of churn.
- High usage of evening minutes, also correlated with higher churn.

Lesser contributors include:

- Number of voicemail messages
- Total international calls
- State
- Area code

Recommendations

- Enhance Customer Service: Reduce wait times and boost overall satisfaction.
- Offer Personalized Call Plans: Provide cost-effective plans tailored for both daytime and nighttime usage.
- Prioritize Service Quality: Continuously monitor and improve network reliability, call quality, and data speed, investing in necessary upgrades.
- Ensure Transparent Pricing: Implement clear pricing structures and billing processes to avoid disputes and enhance satisfaction.
- Engage Proactively with Customers: Regularly gather feedback, address concerns, and offer assistance to prevent churn.
- Strengthen Security Measures: Implement robust protocols to protect voicemail and ensure customer privacy and data security.
- Expand International Plans: Offer diverse international plan options to meet various customer needs.
- Conduct Regular Churn Analysis: Regularly analyze churn patterns to inform proactive retention strategies.



NEXT STEPS

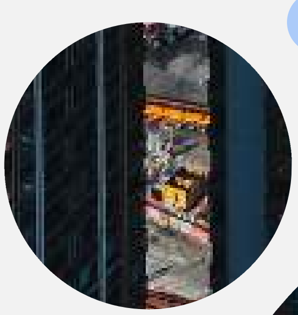
Deploy the Model: Integrate the churn prediction model into the operational system to provide real-time predictions on customer churn, enabling proactive retention strategies.

Monitor and Update the Model: Continuously monitor the model's performance and accuracy, regularly updating it with new data to ensure it effectively predicts churn over time.

Interpret Model Insights: Analyze the model's predictions to identify the main drivers of customer churn, providing valuable insights for targeted retention efforts and strategic decision-making.

Diversify Data: Expand the dataset by collecting a wider range of customer attributes, behaviors, and interactions to enhance the model's predictive capabilities and capture more nuanced churn patterns.

THANK YOU



BRIAN KARIITHI