PREDICTING CUSTOMER CHURN: INSIGHTS AND STRATEGIES FOR SYRIATEL

INTRODUCTION TO CUSTOMER CHURN

This project leverages the first dataset to develop a predictive model aimed at improving customer retention by identifying potential churners.

Objective: Our objective was to build a model that not only achieves high accuracy but also balances precision and recall to ensure reliable predictions.

Importance: By accurately predicting [specific outcome], stakeholders can take specific action or decision-making to foster improvement.

PROJECT GOALS

Primary Goal: To create a robust classification model using the first dataset that can reliably predict customer churn.

Secondary Goal: To ensure the model is interpretable and actionable for stakeholders, enabling data-driven decisions.

Stakeholder Connection: This project supports the marketing department by providing insights that can lead to better personalized marketing strategies.

DATA OVERVIEW

Dataset Overview: We used the first dataset, which includes information about customer call minutes, different subscription plans and charges.

Key Features: The dataset consists of 22 features.

Preprocessing Steps: Handled categorical values.

Standardized/normalized features to ensure consistency.

Applied feature selection techniques to reduce dimensionality and improve model performance.

METHODOLOGY

Model Selection: After experimenting with various algorithms, we selected the Random Forest Classifier due to its superior performance.

Cross-Validation: To ensure the model's reliability, we used cross-validation, which provided consistent and reliable metrics across different subsets of the data.

Hyperparameter Tuning: We optimized the model using grid search, focusing on key parameters like n_estimators, max_depth, min_samples_leaf.

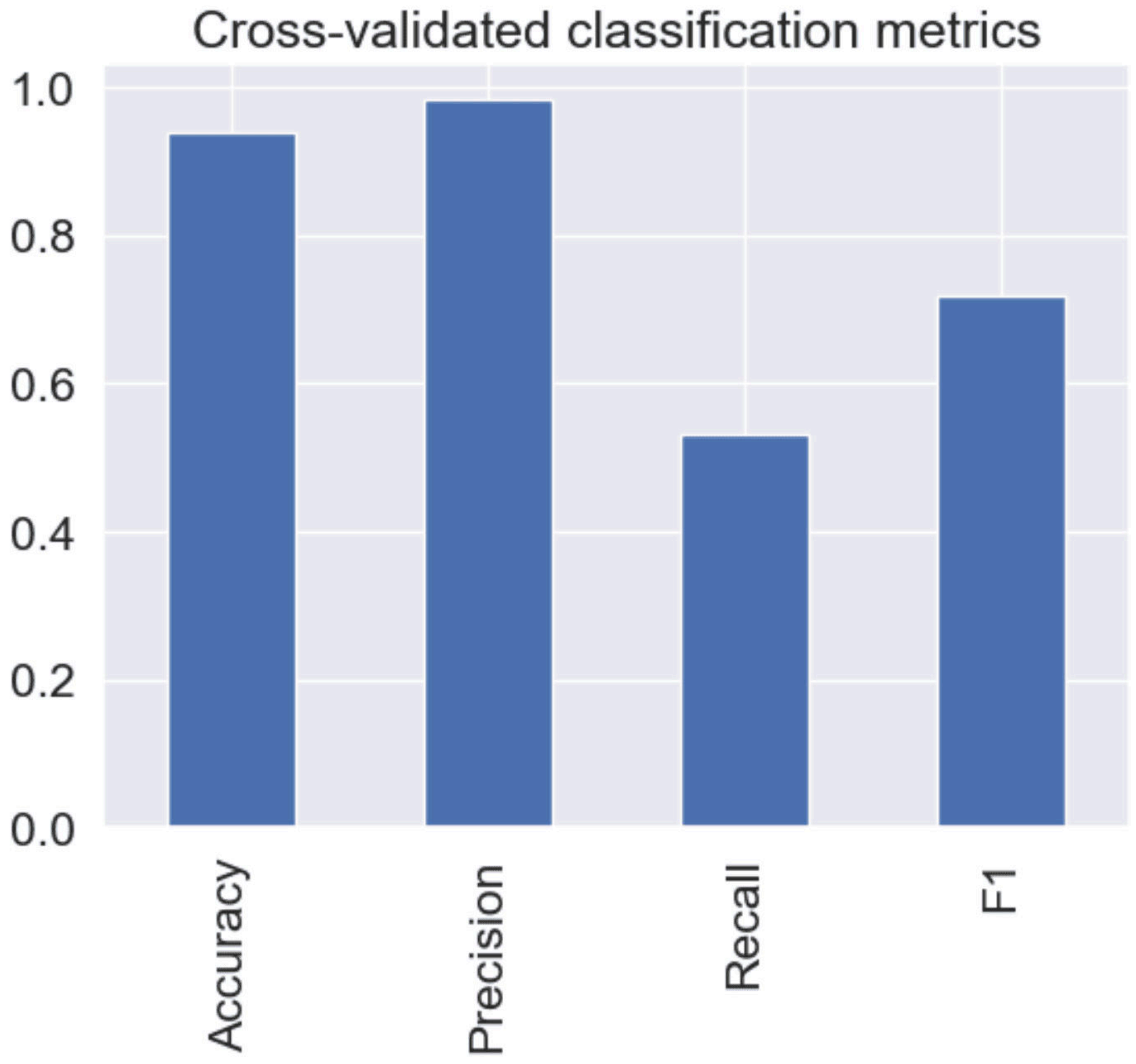
MODEL PERFORMANCE

Accuracy: The model achieved an accuracy of94%, indicating that the majority of predictions were correct, with only a small proportion of errors.

Precision: High precision at 98% reflects the model's ability to minimize false positives, meaning that when the model predicts a positive class, it is almost always correct.

Recall: The recall of 53% suggests that the model was able to correctly identify just over half of the actual positive cases. While this is lower than ideal, it indicates a trade-off where the model is conservative in predicting positives, possibly to avoid false positives.

F1 Score: An F1 score of 72% shows a balanced performance between precision and recall, combining both metrics into a single number that reflects the overall accuracy of the positive predictions.



ROC ANALYSIS

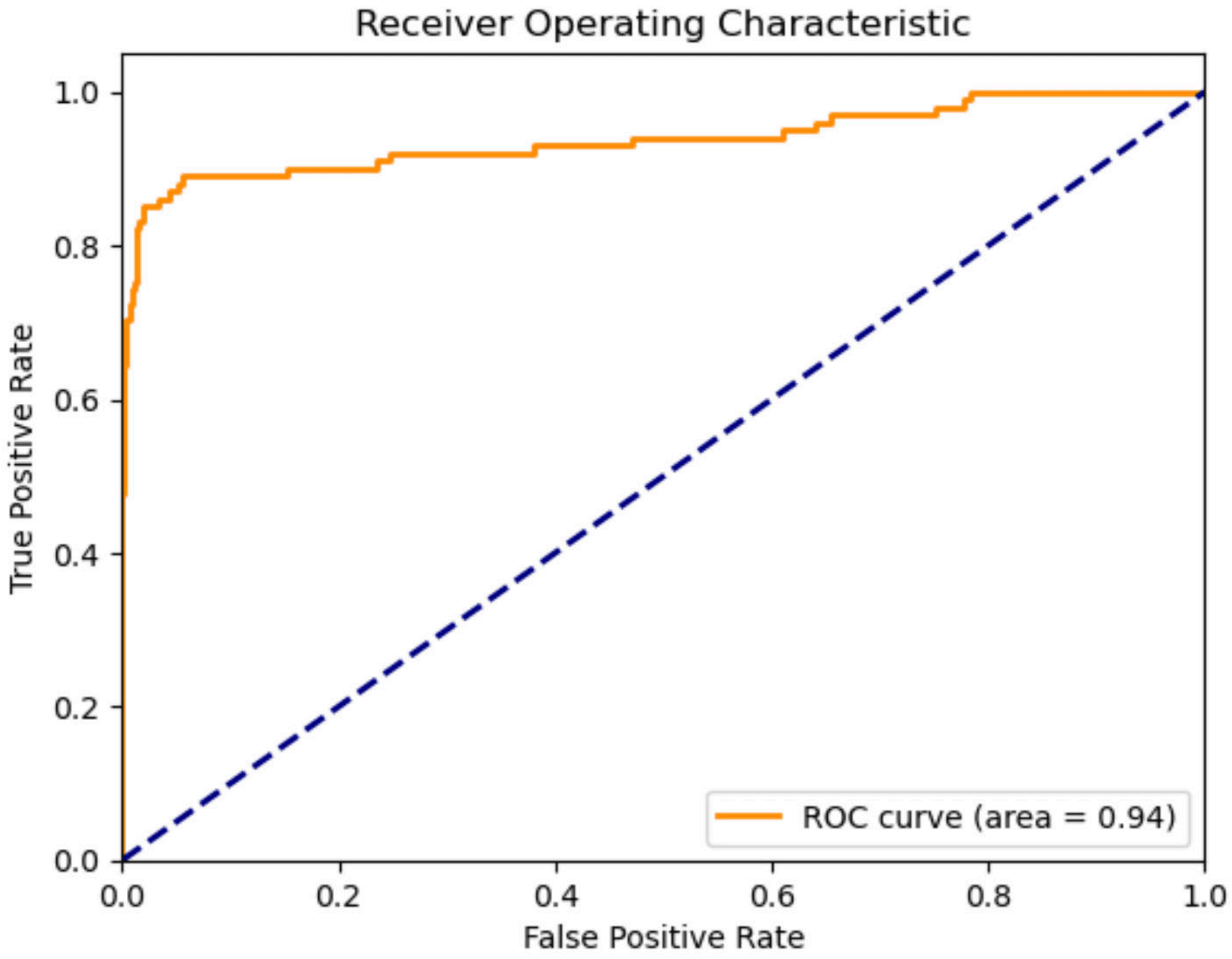
AUC: The area under the curve (AUC) is 0.94, indicating excellent model performance.

True Positive Rate vs. False Positive Rate: The ROC curve shows the trade-off between true positive rate and false positive rate, with our model consistently achieving high true positive rates across thresholds.

MODEL OPTIMIZATION

Optimal Number of Neighbors: The optimal KNN model was achieved yielding a test accuracy of 88.91%.

Overfitting vs. Generalization: The plot shows how increasing the number of neighbors affects the model’s ability to generalize, with the training and test scores converging as the model becomes more robust.



CONCLUSION AND NEXT STEPS

Summary: The project successfully developed a predictive model with strong performance metrics, demonstrating its potential for predicting customer churn.

RECOMMENDATIONS

- Deploy the model for real-time predictions in the marketing department.
- Consider offering incentives for long-term contracts to customers with high churn risk, particularly those with high total minutes.
- Implement proactive customer service initiatives aimed at addressing issues before they lead to churn.

Thanks!

Do you have any questions?