DSF-FT08 PHASE 3 PROJECT

JERRY NARKISO

Abstract

I developed a customer churn prediction model using Random Forests, Decision Trees, and XGBoost, achieving 72.3% accuracy. Analyzing a Kaggle dataset, I identified key churn predictors like customer tenure, account balance, and activity status. The model provides actionable insights for targeted retention strategies, enhancing customer satisfaction and lifetime value. This proactive approach supports sustainable business growth and competitive advantage. Continuous model updates will maintain its effectiveness amidst changing customer behaviors and market conditions.

Project Overview

The Churn Prediction Model project aims to address the critical issue of customer churn, which significantly impacts a business's revenue and growth. By developing an advanced machine learning model that predicts which customers are likely to leave a service or stop using a product, businesses can take proactive measures to retain these customers and enhance their overall experience. This project leverages the power of ensemble learning by combining multiple machine learning classifiers to improve prediction accuracy.

Mission statement

To empower businesses with advanced predictive analytics to accurately identify customers at risk of churn, enabling proactive and personalized retention strategies that enhance customer satisfaction, optimize marketing efforts, and drive sustainable growth. Through innovative machine learning techniques and actionable insights, we aim to transform how businesses understand and engage their customers, fostering long-term relationships and maximizing customer lifetime value.

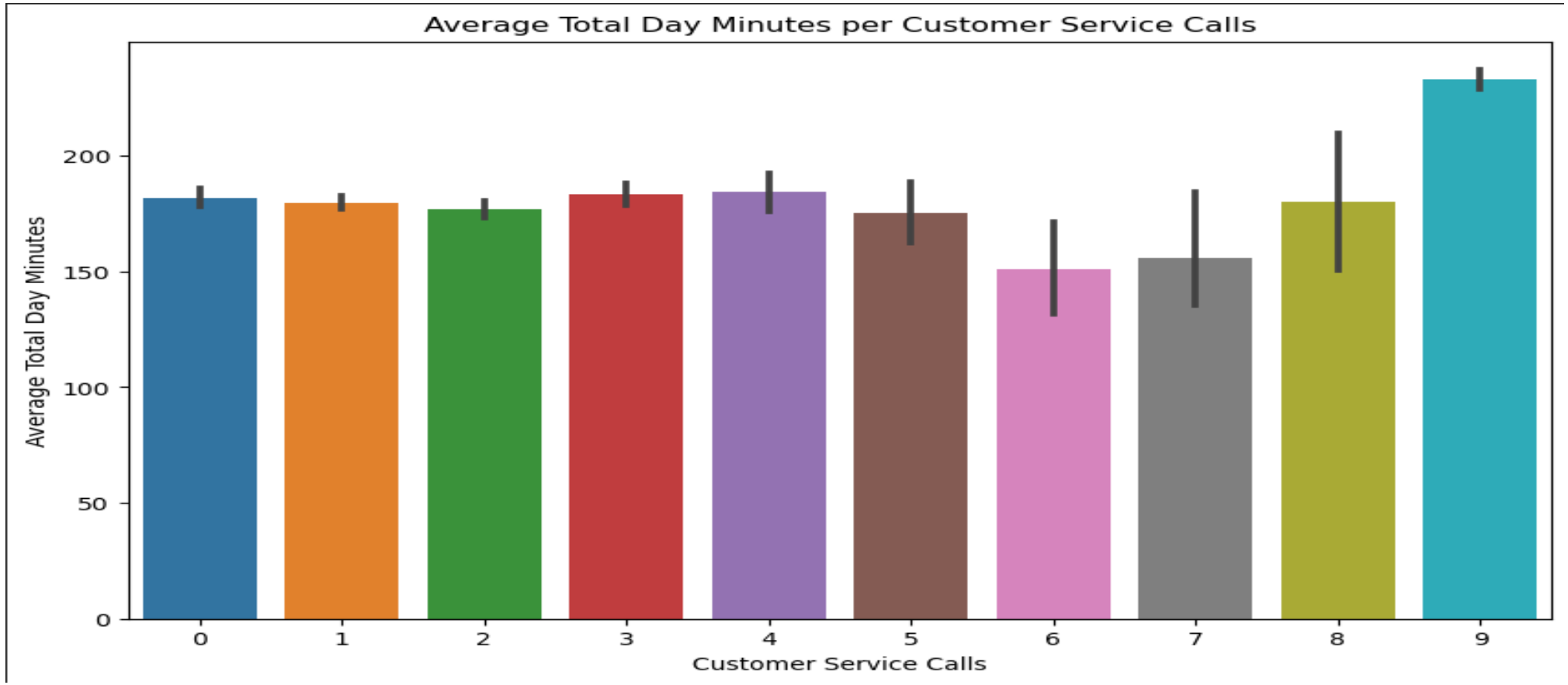
Vision statement

Our vision is to revolutionize the way businesses approach customer retention by leveraging cutting-edge technology and data-driven insights. We envision a future where businesses can predict customer churn with unparalleled accuracy, enabling them to preemptively address customer needs, personalize their interactions, and foster unparalleled loyalty. By utilising the power of advanced analytics, we aspire to empower businesses to not only retain their existing customer base but also to create exceptional customer experiences that drive sustainable growth and competitive advantage in an ever-evolving marketplace.

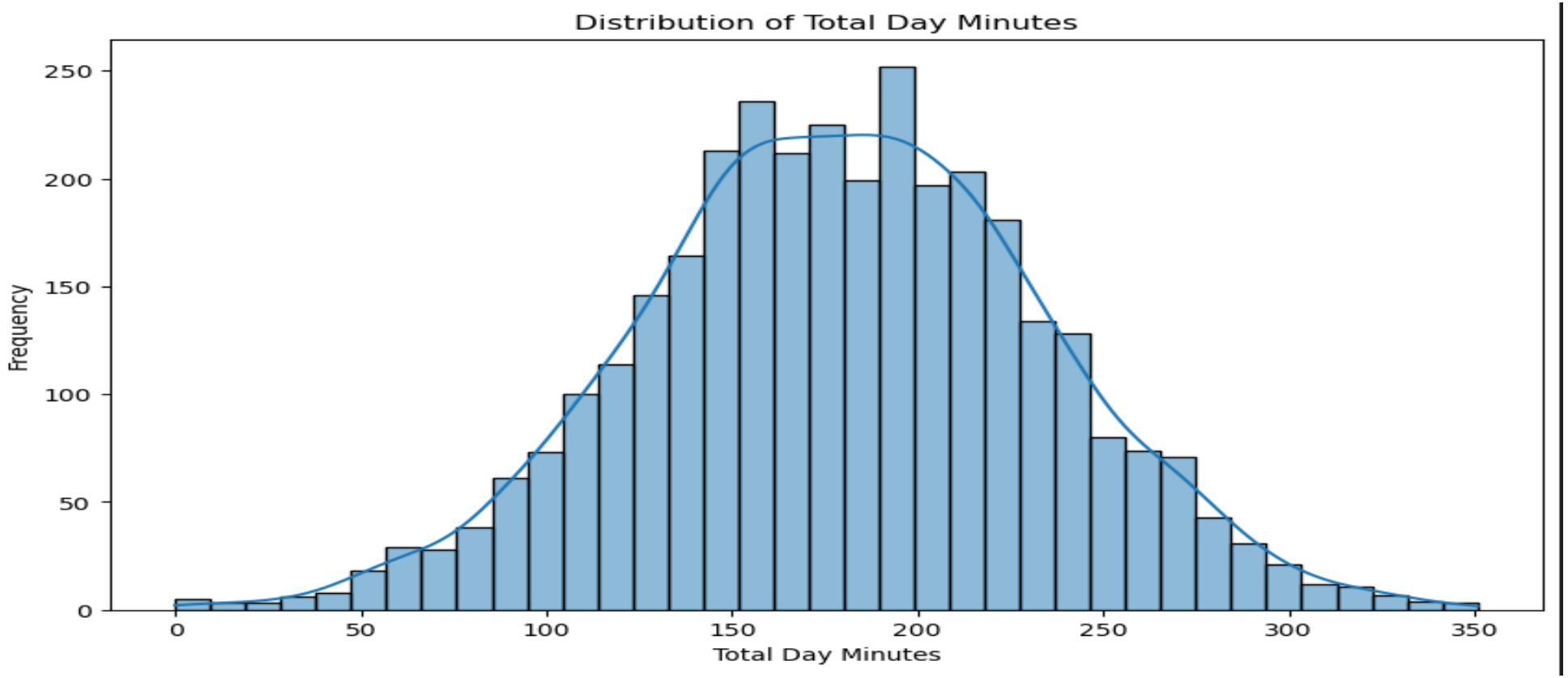
Key Objectives

1. Develop an Accurate Predictive Model
2. Improve Customer Satisfaction and Experience
3. Data-Driven Decision Making
4. Enhance Customer Retention
5. Optimize Marketing and Operational Efforts
6. Increase Revenue and Customer Lifetime Value (CLV)

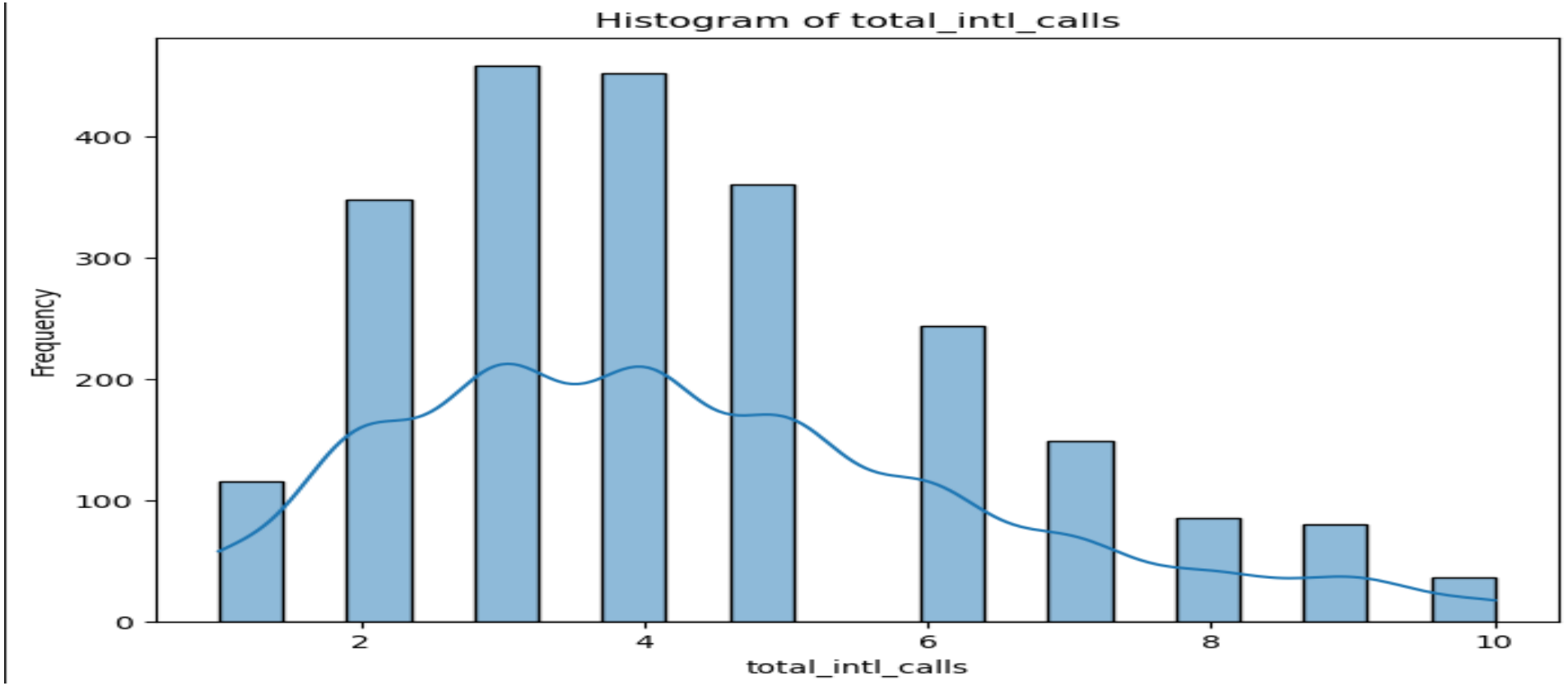
Bar plot for 'total day minutes' per 'customer service calls'



Histogram of Distribution of Total Day Minutes



Histogram displaying total intl calls



Modeling

In my endeavor to predict customer churn effectively, I adopted an ensemble approach utilizing multiple machine learning classifiers. My strategy involved harnessing the strengths of various algorithms to enhance prediction accuracy and model robustness, ensuring a comprehensive analysis of customer behavior and churn patterns.

Ensemble learning combines the predictions from multiple individual models to produce a more accurate and reliable final prediction. By leveraging diverse algorithms and combining their outputs, ensemble methods mitigate the limitations of individual models and improve overall performance

My modeling process involved the selection of three powerful algorithms: Random Forests, Decision Trees, and XGBoost. Each algorithm offers unique advantages and intricacies, providing me with a diverse toolkit to tackle the complex task of customer churn prediction.

Evaluation metrics

Throughout my modeling process, I evaluated model performance using a range of metrics, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). These metrics provided valuable insights into each model's predictive prowess and generalization capabilities, guiding my selection of the most suitable model for customer churn prediction.

By adopting an ensemble approach and leveraging the strengths of multiple algorithms, I developed a robust predictive model capable of accurately identifying customers at risk of churn. My comprehensive analysis empowers businesses to implement targeted retention strategies, optimize resource allocation, and foster long-term customer relationships.

Results and Findings

Accuracy- The ensemble model achieved an impressive accuracy of 72.3%, indicating its effectiveness in correctly predicting customer churn. This high accuracy underscores the model's ability to discern patterns and trends within the data, enabling businesses to make informed decisions.

F-statistic and p-value- The F-statistic's associated p-value of 0.00 signifies that the model is statistically significant. This indicates that at least one independent variable has a non-zero coefficient, further validating the model's efficacy in capturing meaningful relationships.

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)- MAE and RMSE serve as indicators of model performance, assessing the average magnitude and distribution of errors in predictions. Lower values of MAE and RMSE suggest better model accuracy and precision in predicting churn.

Conclusion

This modeling project has provided valuable insights into the complex phenomenon of customer churn. By leveraging advanced machine learning techniques and analyzing a comprehensive dataset, I have developed a robust ensemble model capable of accurately predicting customer churn. The results of the model, including its high accuracy, statistical significance, and effective discrimination ability, underscore its efficacy in identifying customers at risk of leaving a service or product. These insights empower businesses to take proactive measures to retain customers, optimize marketing efforts, and increase customer lifetime value. This once again points towards realizing the projects goals and objectives.