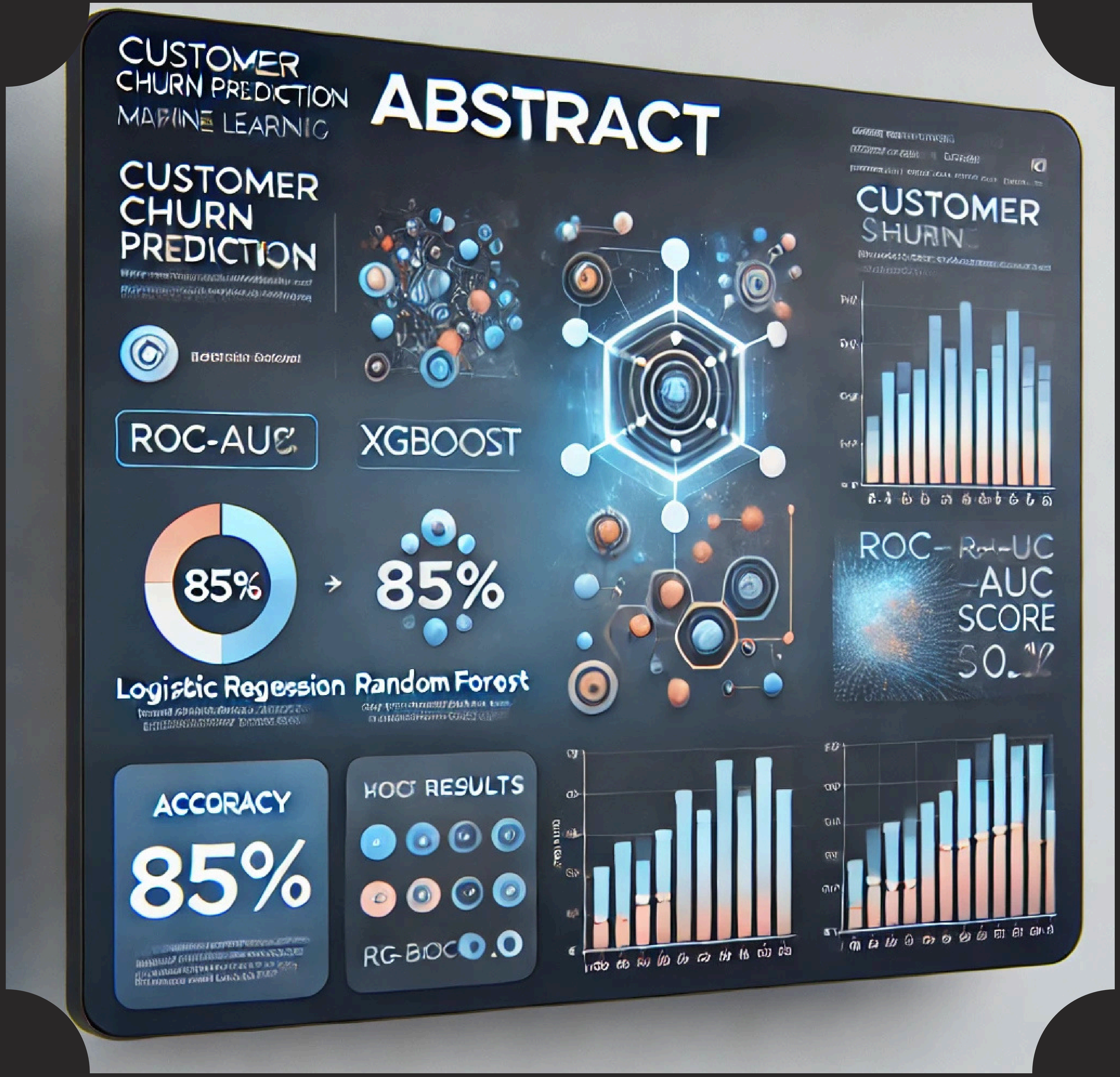
CUSTOMER CHURN PREDICTION USING MACHINE LEARNING

By Sahitha Koppula,
Niharika Goud Cika,
Archana Srinivas

ABSTRACT

- Customer churn is a major issue in industries like telecom, where acquiring new customers is costlier than retaining existing ones.
- This project uses machine learning to predict at-risk customers and recommend strategies to retain them.
- **<u>Models Used</u>**: Logistic Regression, Random Forest, XGBoost, and Neural Networks.
- **<u>Key Results</u>**: XGBoost achieved the highest accuracy of 85% and a ROC-AUC score of 0.93.



INTRODUCTION

- What is Customer Churn?
  Customers leave or stop using services.
  Example: Switching to competitors due to better pricing.

- Why It Matters?
  Acquiring new customers costs 5x more than retaining existing ones.Churn directly impacts revenue.

- Goal: Predict churn early to take action.

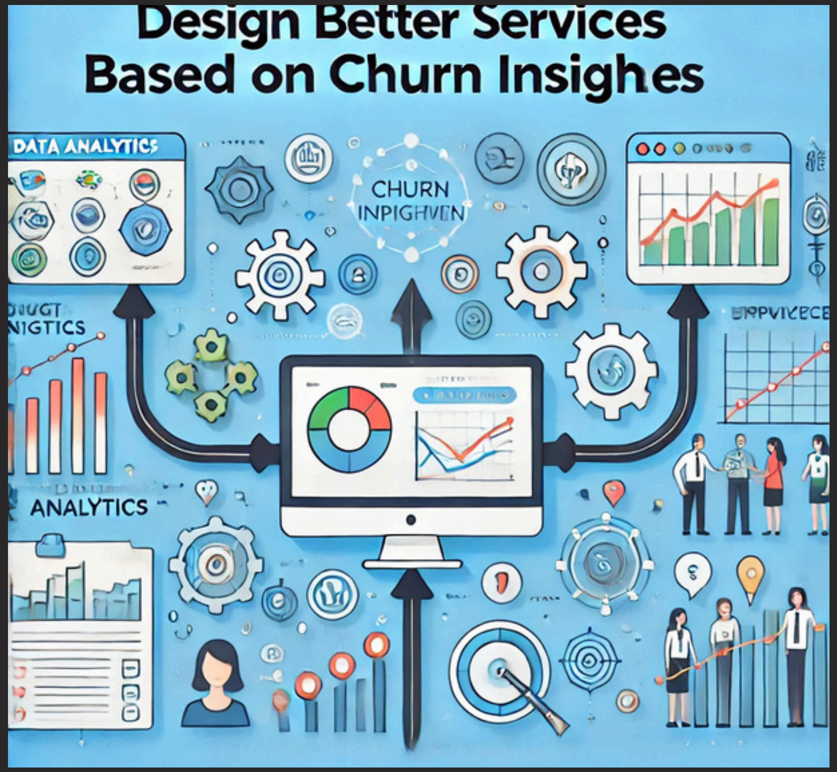
HOW MACHINE LEARNING HELPS

Why Machine Learning?
Processes large amounts of data efficiently.
Identifies patterns and predicts churn with high accuracy.

APPLICATIONS

Offer discounts or promotions to at-risk customers.

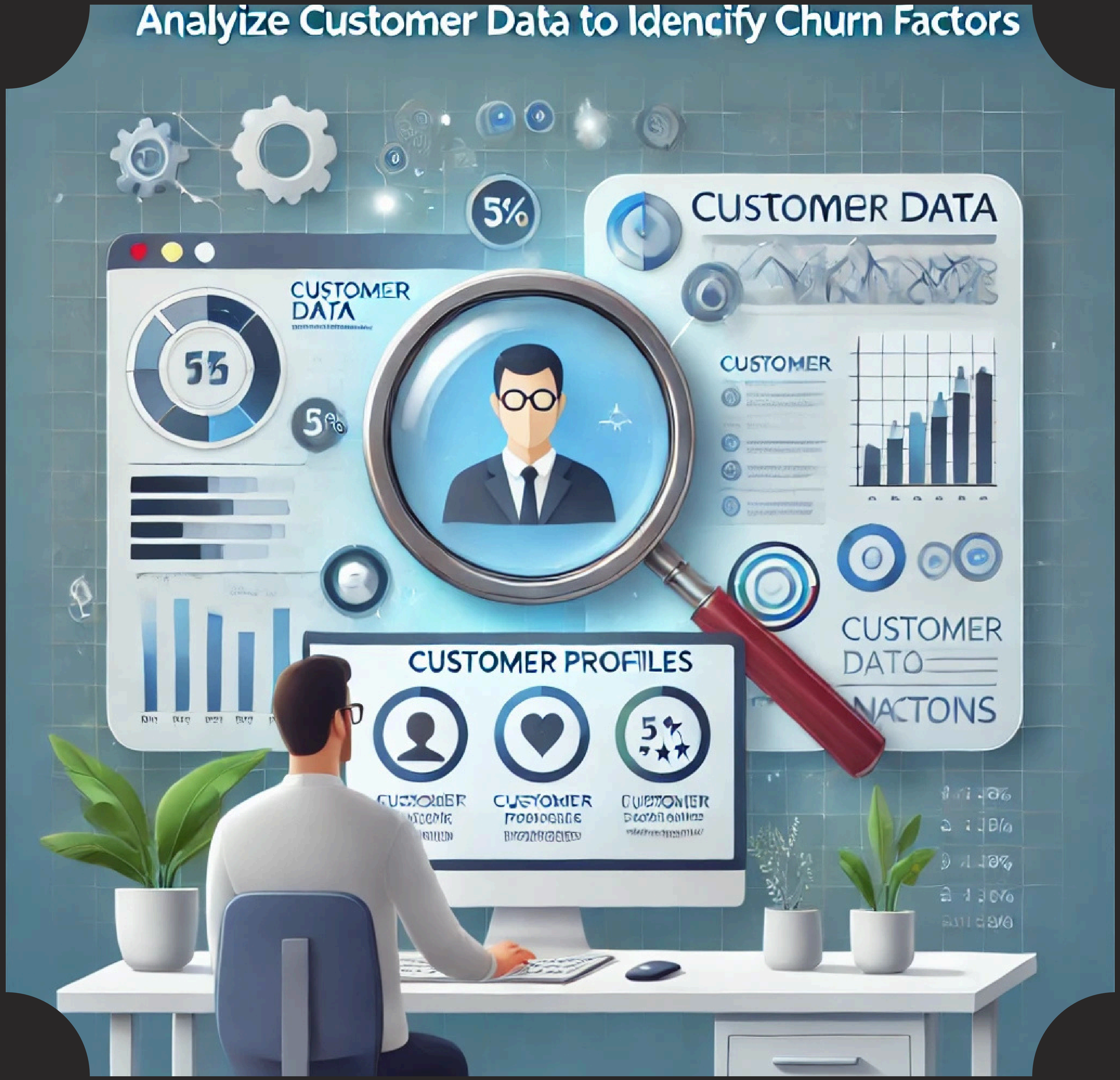
Improve customer support for specific groups.



Design better services based on churn insights.

OBJECTIVES

- Analyze customer data to identify churn factors.
- Implement and compare machine learning models.
- Evaluate performance metrics: Accuracy, Recall, ROC-AUC.
- Recommend strategies for real-world deployment.

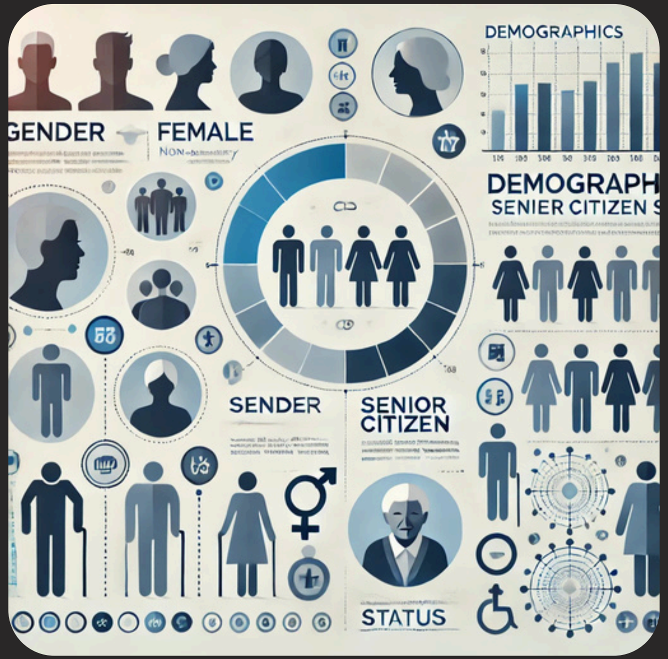


DATASET OVERVIEW

Source: Telco Customer Churn Dataset (Kaggle).

Size: 7,043 customer records.

Features



Demographics: Gender, senior citizen status.

Account Info: Contract type, payment method.

GService Details: Internet, streaming, tech support. rayson Burch

Financials: Monthly charges, total charges.

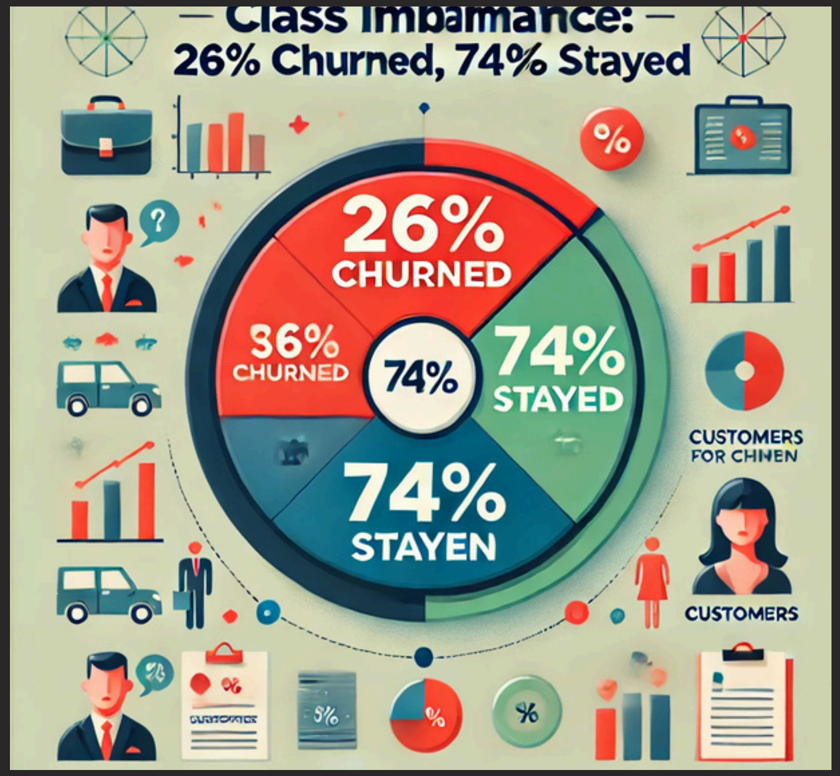
Target: Churn (Yes/No).

PREPROCESSING STEPS:

- Filled missing values with median.
- Encoded categorical variables (e.g., one-hot encoding).
- Scaled numerical features with Min-Max Scaling.
- Used SMOTE to address class imbalance.

CHALLENGES IN DATASET

Missing Values: Found in TotalCharges.



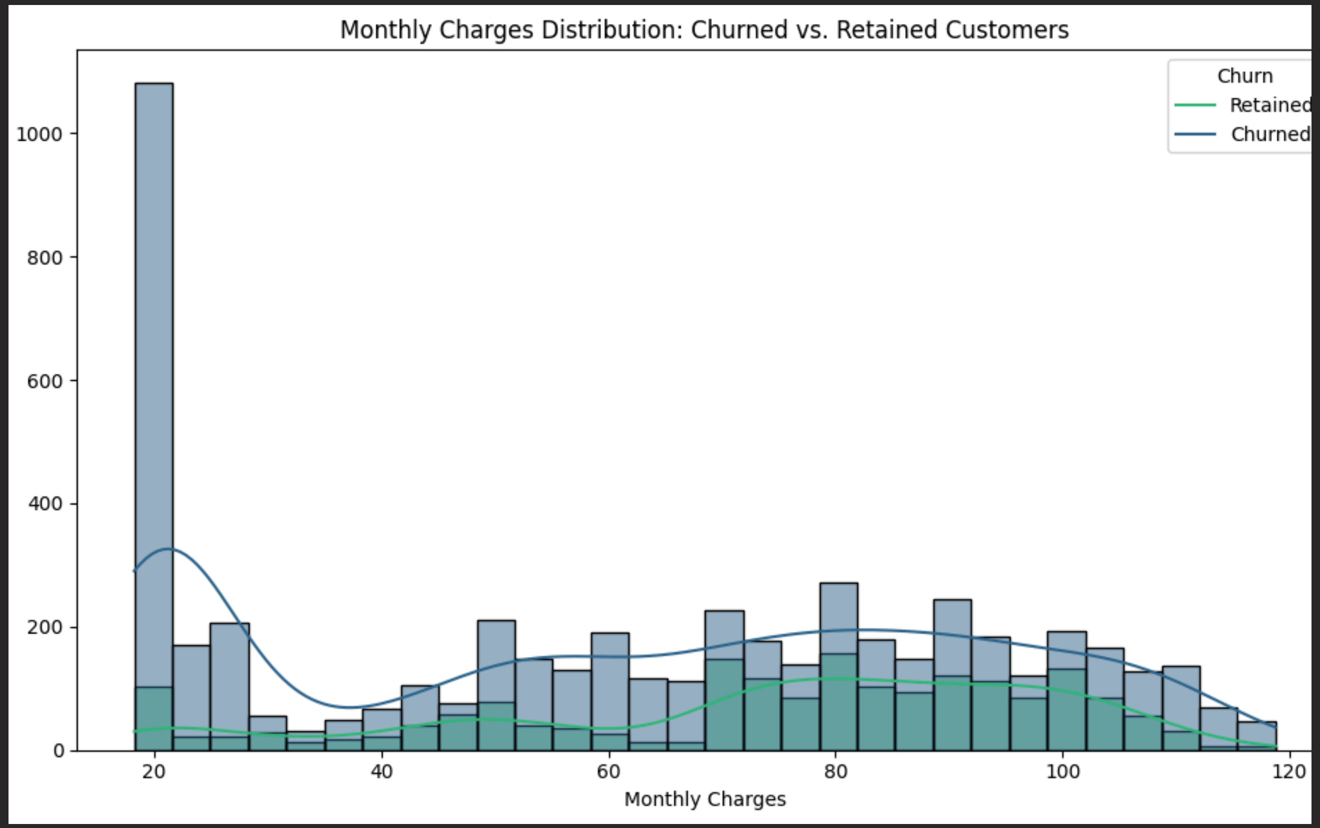
Class Imbalance: 26% churned, 74% stayed.

Categorical Data: Non-numerical columns (e.g., Contract).

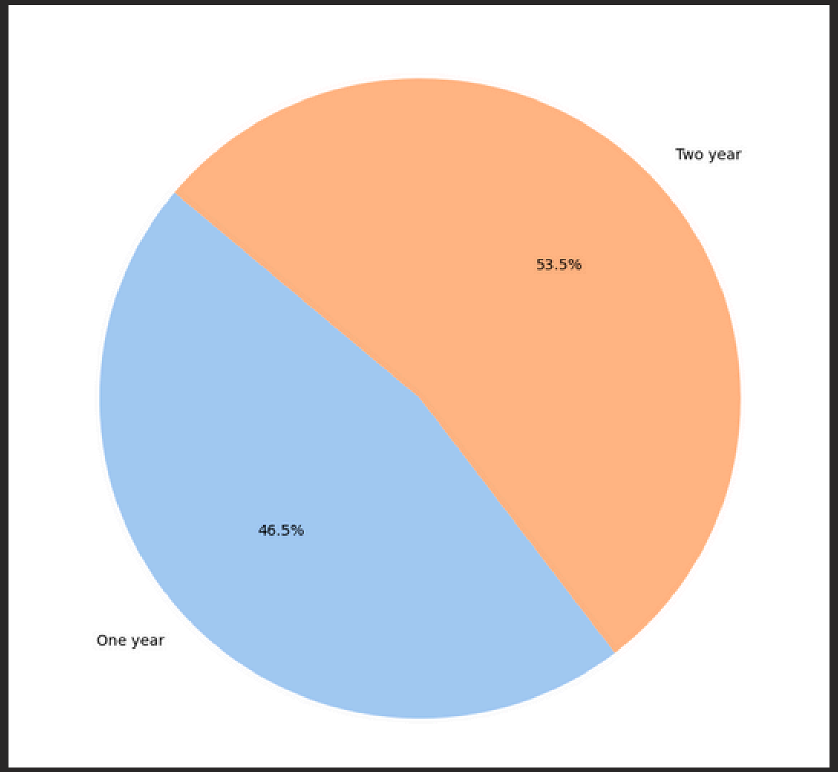
EXPLORATORY DATA ANALYSIS

- **Key Insights from EDA**:
- **Churn Rate**: 26% customers churned.
- **Tenure**: Shorter tenure correlates with churn.
- **Monthly Charges**: Higher charges increase churn risk.

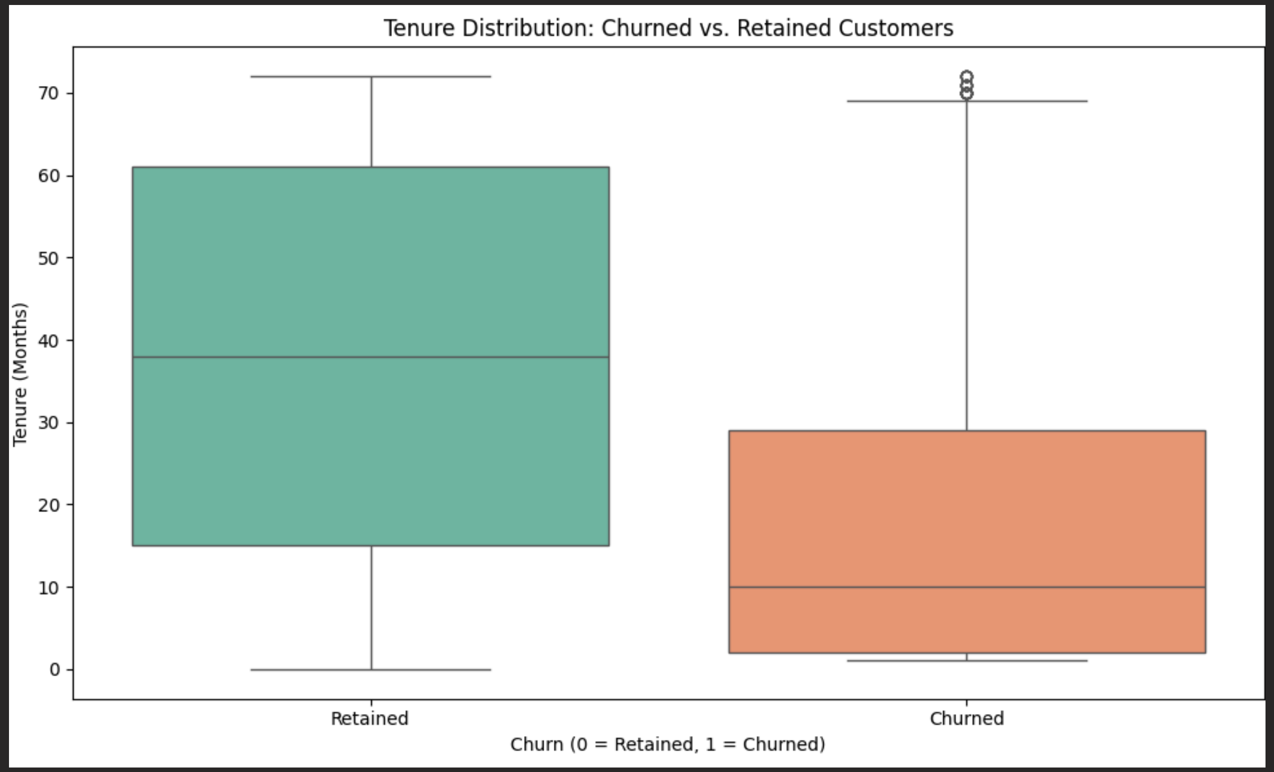
VISUALIZATIONS



Histogram: High charges linked to churn.

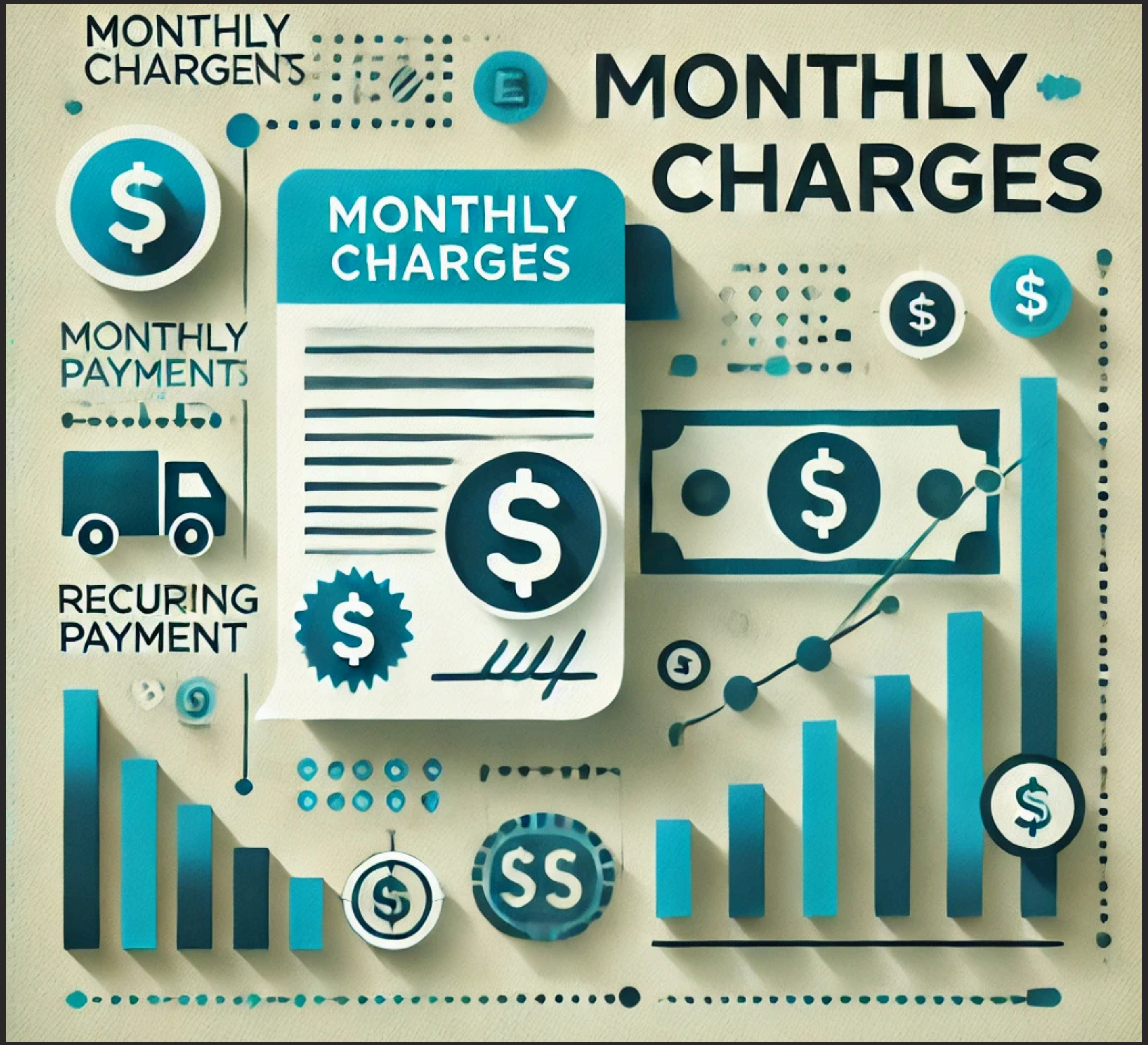


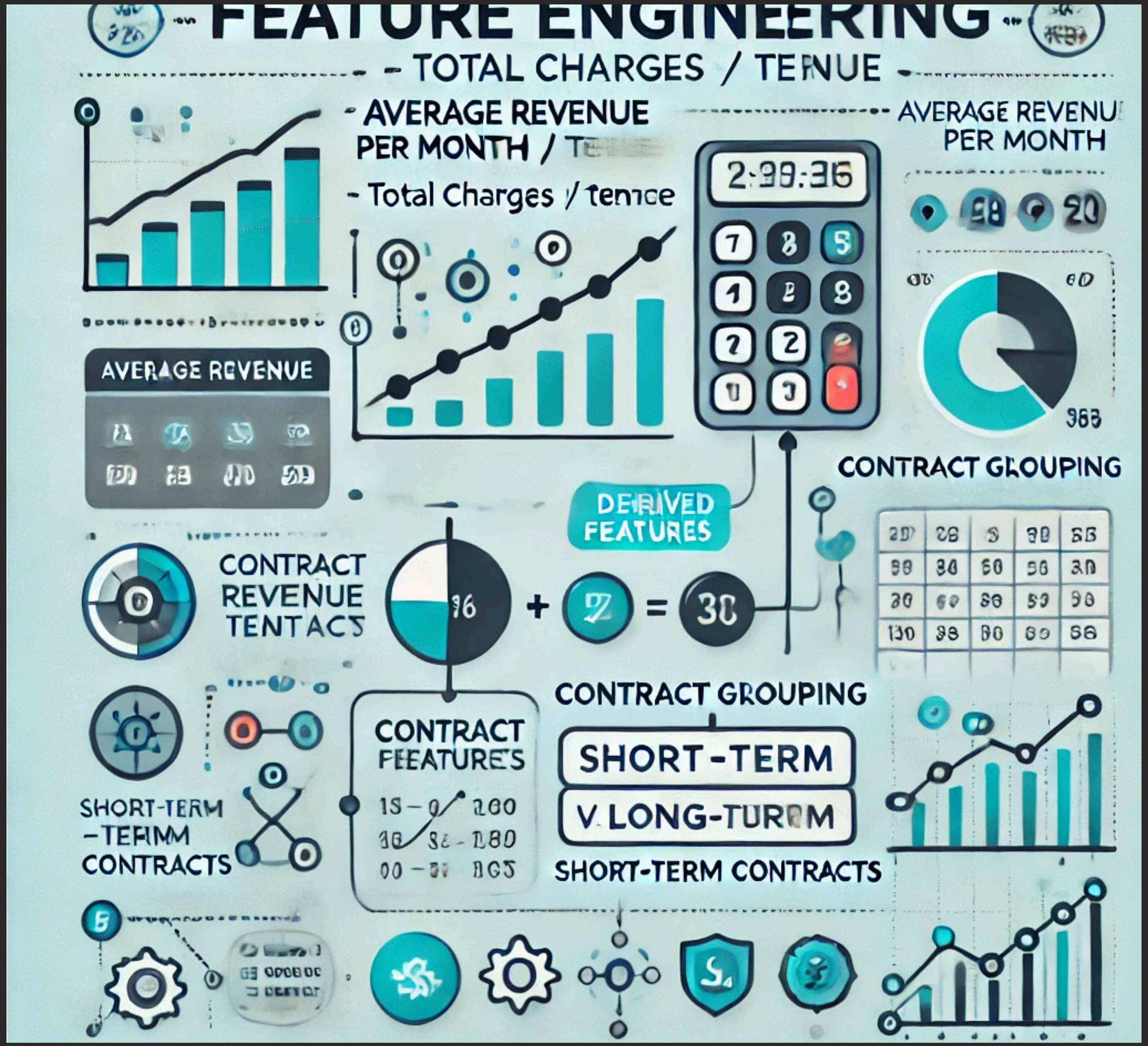
Pie Chart: Most churned customers had month-to-month contracts.



Box Plot: Shorter tenure customers more likely to churn.

FEATURE ENGINEERING





<u>Selected Features</u>:

- Tenure.
- Monthly Charges.
- Contract Type.
- Payment Method.

<u>Derived Features</u>:

Average Revenue Per Month = TotalCharges / Tenure.

Contract Grouping: Short-term vs. Long-term contracts.

MODELS IMPLEMENTED

Logistic Regression:

Pro: Simple, interpretable baseline.

Con: Limited to linear relationships.

Random Forest:

Pro: Handles non-linear data, shows feature importance.

Con: Computationally expensive.

XGBoost:

Pro: Effective for imbalanced datasets.

Con: Requires hyperparameter tuning.

Neural Networks:

Pro: Captures complex relationships.

Con: Resource-intensive, hard to interpret.

METHODOLOGY

Steps:

Split the dataset into 70% training and 30% testing.
Applied SMOTE to balance classes.
Used 5-fold cross-validation for generalization.
Tuned hyperparameters for Random Forest and XGBoost.

Evaluation Metrics:

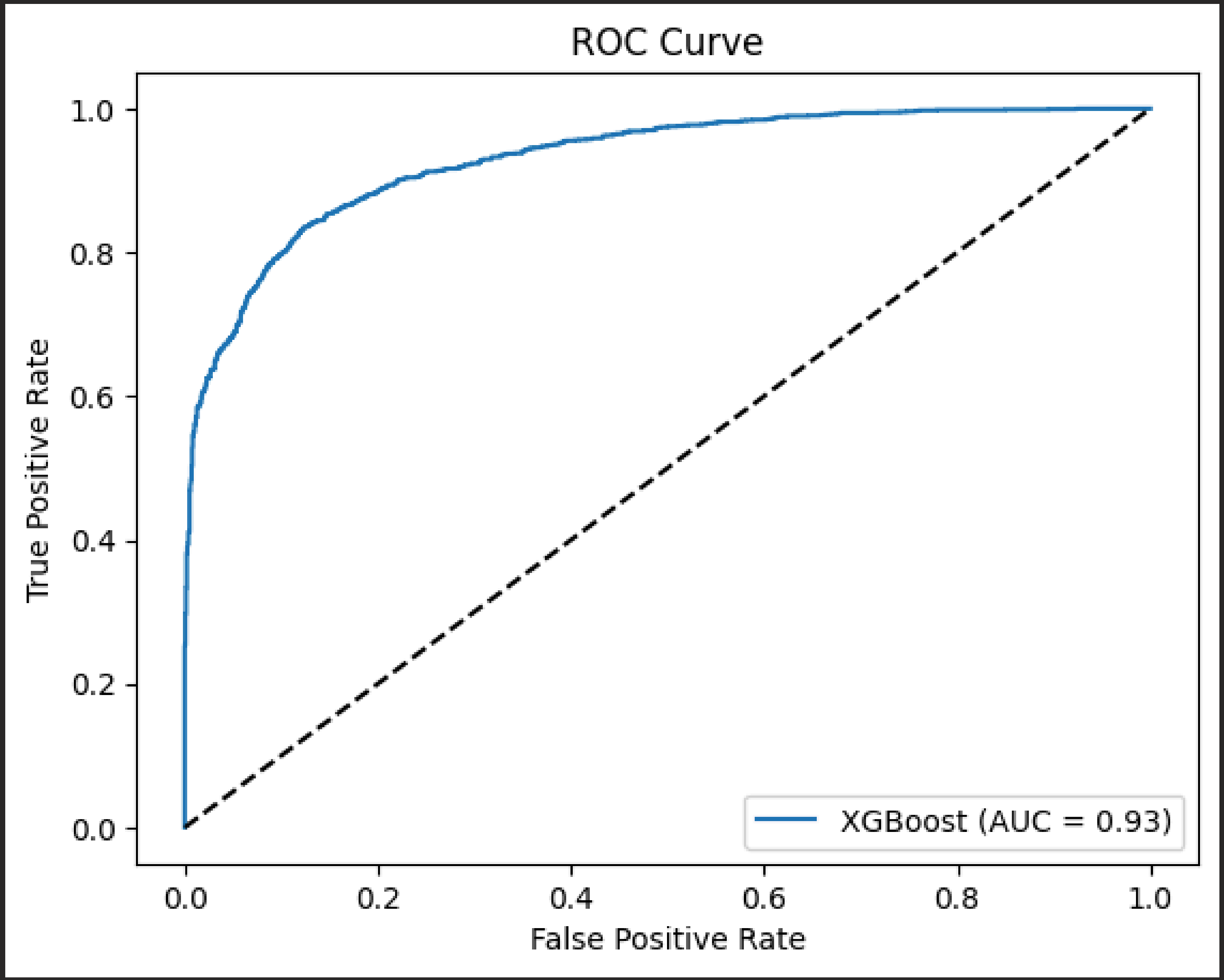
Accuracy, Precision, Recall, F1-Score, ROC-AUC.

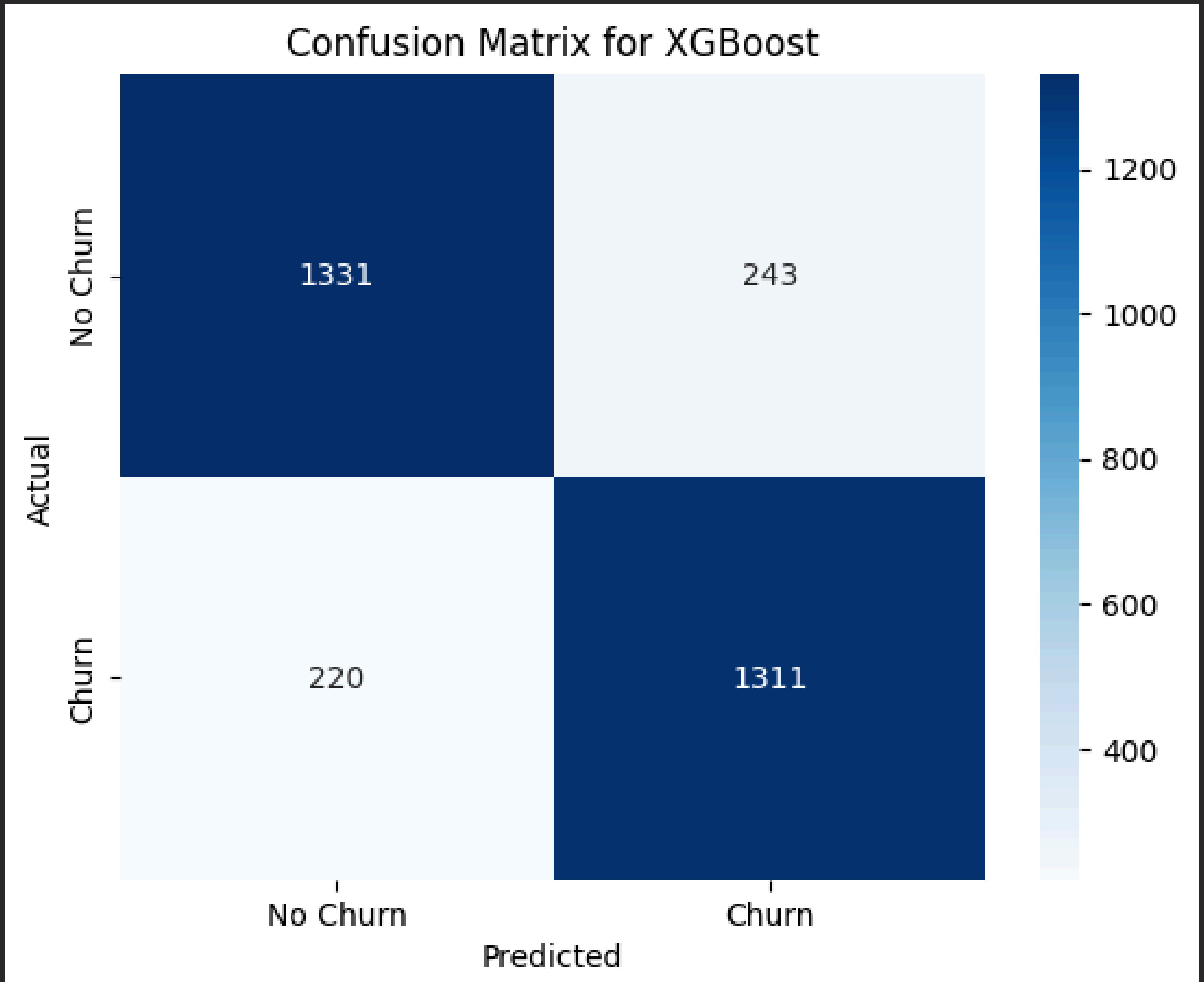
RESULTS

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	81%	80%	83%	81%	0.9
Random Forest	85%	85%	84%	84%	0.92
XGBoost	85%	84%	85%	84%	0.93
Neural Networks	82%	80%	85%	82%	0.91

VISUAL REPRESENTAION





KEY INSIGHTS

Best Model: XGBoost

Highest Accuracy (85%) and ROC-AUC (0.93).

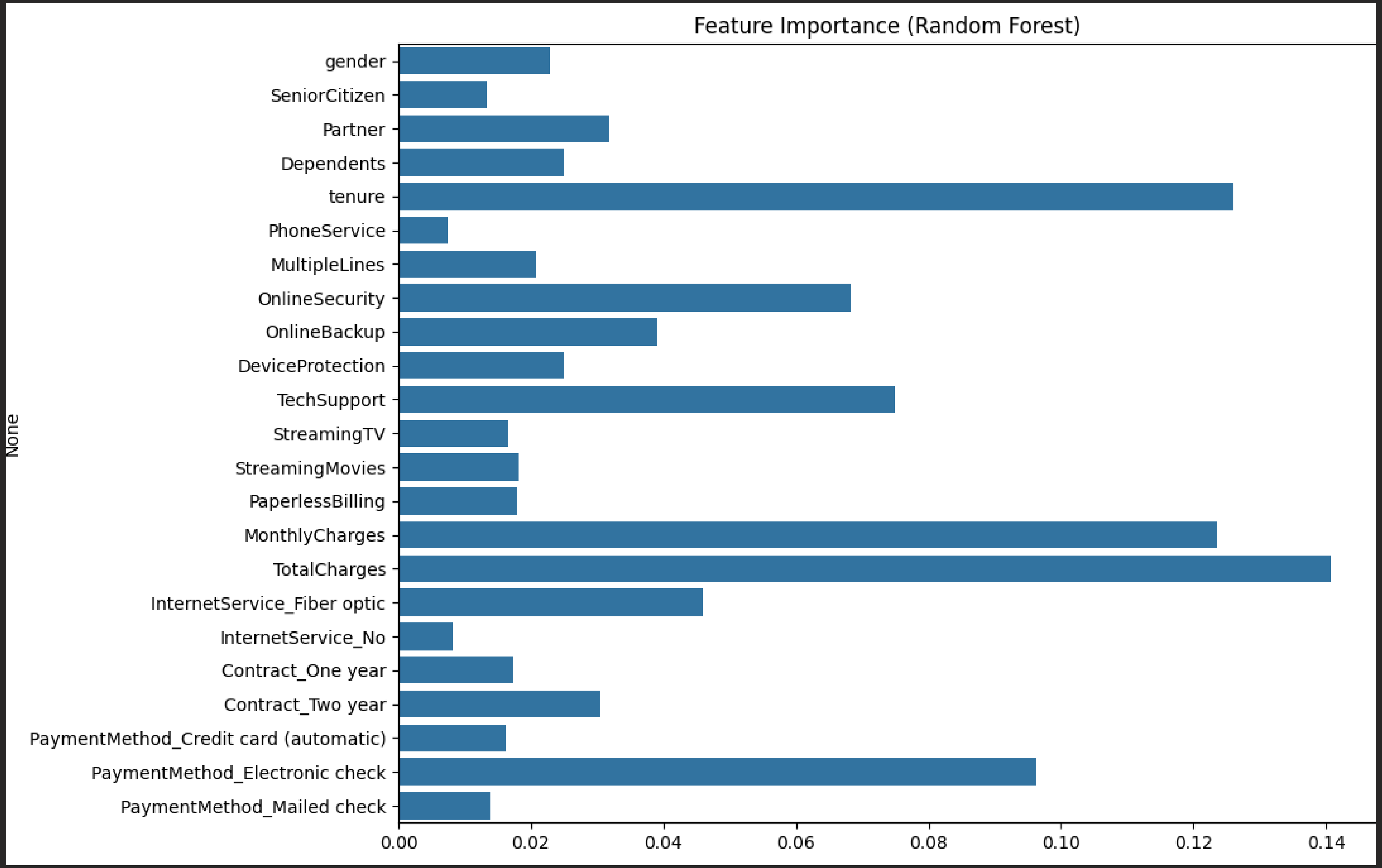
Feature Importance:

Tenure, Monthly Charges, and Contract Type are the top predictors.

Trade-offs:

Logistic Regression: Interpretable but less accurate.

Neural Networks: Competitive but resource-intensive.



RECOMMENDATIONS

Improvement Areas:

Incorporate customer feedback and sentiment analysis.

Experiment with ensemble methods combining model strengths.

Deployment:

Integrate XGBoost into CRM systems for real-time churn predictions.

Use insights to proactively engage at-risk customers.

REFERENNCES

1. Adwan, Omar, et al. “Predicting Customer Churn in Telecom Industry Using Multilayer Perceptron Neural Networks: Modeling and Analysis.” Life Science Journal, vol. 11, no. 3, 29 Jan. 2014, pp. 75–81, www.researchgate.net/publication/260944174_Predicting_Customer_Churn_in_Telecom_Industry_using_Multilayer_Preceptron_Neural_Networks_Modeling_and_Analysis
2. “IEEE Xplore Full-Text PDF”: Ieee.org, 2024, ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10531735
3. Sam, Glory, et al. “Customer Churn Prediction Using Machine Learning Models.” Journal of Engineering Research and Reports, vol. 26, no. 2, 6 Feb. 2024, pp. 181–193, https://doi.org/10.9734/jerr/2024/v26i21081 Accessed 20 October 2024
4. Das, Debendra, and Sheetal Mahendher. “COMPARATIVE ANALYSIS of MACHINE LEARNING APPROACHES in PREDICTING TELECOM CUSTOMER CHURN.” Educational Administration Theory and Practice Journal, vol. 30, no. 5, 24 May 2024, pp. 8185–8199, www.researchgate.net/publication/380823708 , https://doi.org/10.53555/kuey.v30i5.4348 Accessed 20 October 2024

REFERENNCES

5. Bogaert, Matthias, and Lex Delaere. “Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-The-Art.” Mathematics, vol. 11, no. 5, 24 Feb. 2023, p. 1137, https://doi.org/10.3390/math11051137 Accessed 20 October 2024

6. Sweidan, Dirar, et al. Predicting Customer Churn in Retailing. 2022, www.diva-portal.org/smash/get/diva2:1727230/FULLTEXT01.pdf

7. Ahmad, Abdelrahim Kasem, et al. “Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform.” Journal of Big Data, vol. 6, no. 1, Mar. 2019, https://doi.org/10.1186/s40537-019-0191-6

8. Rajendran, Srinivasan & Devarajan, Rajeswari & Elangovan, G.. (2023). Customer Churn Prediction Using Machine Learning Approaches. 1-6. 10.1109/ICECONF57129.2023.10083813. (PDF) Customer Churn Prediction Using Machine Learning Approaches


THANK YOU!