SyriaTel Customer Churn Prediction

By Tinah Ngei

Business Objective

- The primary goal of this project is to **predict customer churn** for SyriaTel using historical customer data.
- By identifying patterns and key indicators of churn, we aim to:
- Understand **why customers leave**
- Build accurate **machine learning models** to predict churn
- Provide **data-driven recommendations** to improve customer retention and reduce business losses

Business Understanding

Problem Statement:

- SyriaTel seeks to reduce customer churn. Understanding why customers leave can help retain high-value users.

Goal:

- Build a machine learning model to predict churn and guide business decisions.

Key Business Questions

- Can we predict which customers are likely to churn based on historical data?
- What are the **most important factors** influencing customer churn?
- How can we **reduce churn rates** using actionable insights from the model?
- Which customer segments are most vulnerable to churn, and how should they be addressed differently?

Data Understanding

Data Source: SyriaTel telecom dataset

Key Features:

- \- Customer demographics
- \- Service usage
- \- Billing & payment
- \- Churn status (target)

Total Records: ~7,000+Features: ~20

Tools and Technologies used

- <u>**Data Processing & Analysis**</u>
- **Pandas** – for data wrangling & manipulation
- **NumPy** – numerical operations
- **Matplotlib & Seaborn** – data visualization
- **Scikit-learn** – preprocessing, modeling, and evaluation
- **XGBoost** – for gradient boosting classification
- <u>**Machine Learning Models**</u>
- **Logistic Regression**
- **K-Nearest Neighbors (KNN)**
- **Support Vector Machine (SVM)**
- **Random Forest**
- **XGBoost** (Best Performing Model)
- <u>**Environment & Tools**</u>
- **Jupyter Notebook** – interactive development
- **VS Code** – code editing and version control
- **Git & GitHub** – versioning and collaboration

Data Cleaning & Preparation

Actions Taken:

- - Dropped duplicates
- - Handled missing values
- - Encoded categorical features
- - Scaled numeric values

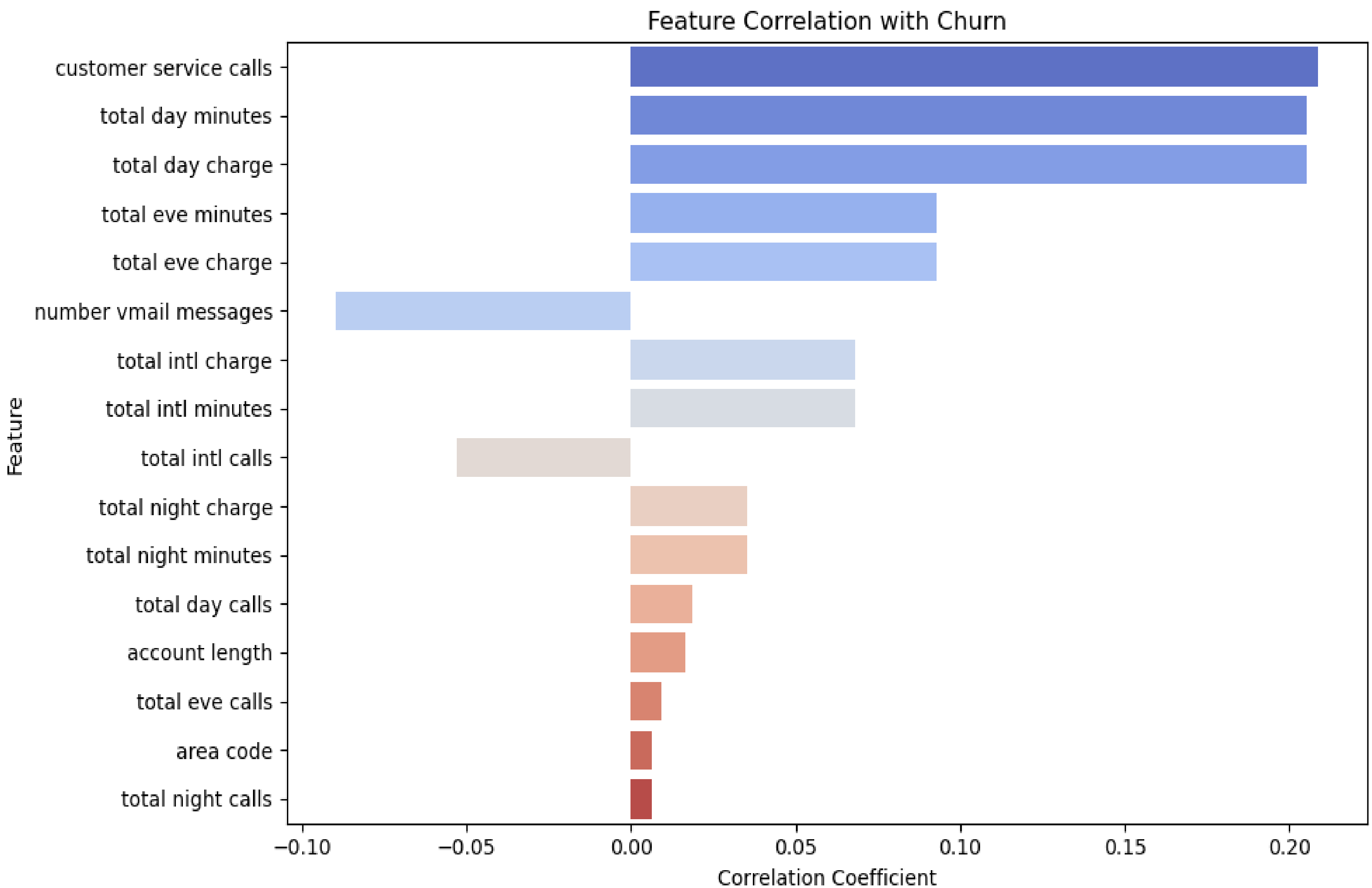
Final Shape: 7032 x 20

Exploratory Data Analysis (EDA)

Key Insights:

- - High churn among Fiber Optic users
- - Longer contracts reduce churn
- - Monthly contracts are high risk

Feature Correlation with Customer Churn



Insights on Feature Correlation with Customer Churn

- **Customer service calls** show the strongest positive correlation (~0.20) with churn, highlighting service interaction as a key churn driver.

- **Daytime usage** features (total minutes/charges) also have a moderate positive impact on churn behaviour.

- **Night usage, area code, and account length** are slightly negatively correlated, suggesting less influence or possible retention signals.

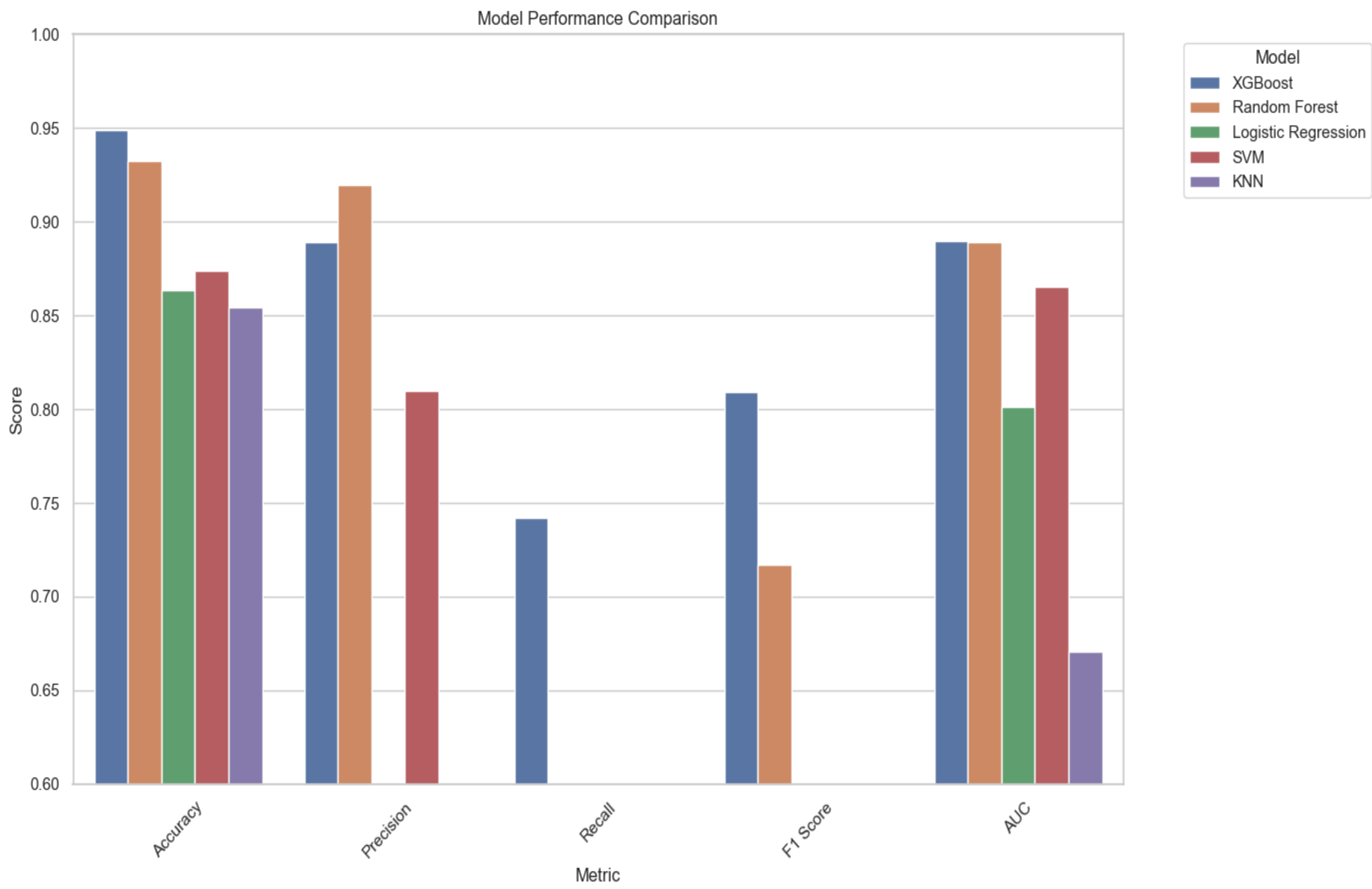
- **Recommendation**: Focus predictive efforts on service call patterns and daytime activity to enhance churn detection and prevention strategies.

Evaluation Metrics

Metrics:

- - Accuracy
- - Precision
- - Recall
- - F1 Score
- - AUC (ROC)

Model Comparison Chart



Recommendation Based on Model Performance Comparison

•**XGBoost** consistently outperforms across **Accuracy, Recall, F1 Score, and AUC**, indicating high reliability in both precision and generalization.

•**Random Forest** leads in **Precision**, making it a strong contender for applications prioritizing true positive rate.

•Other models like **Logistic Regression, SVM, and KNN** show comparatively weaker performance across most metrics.

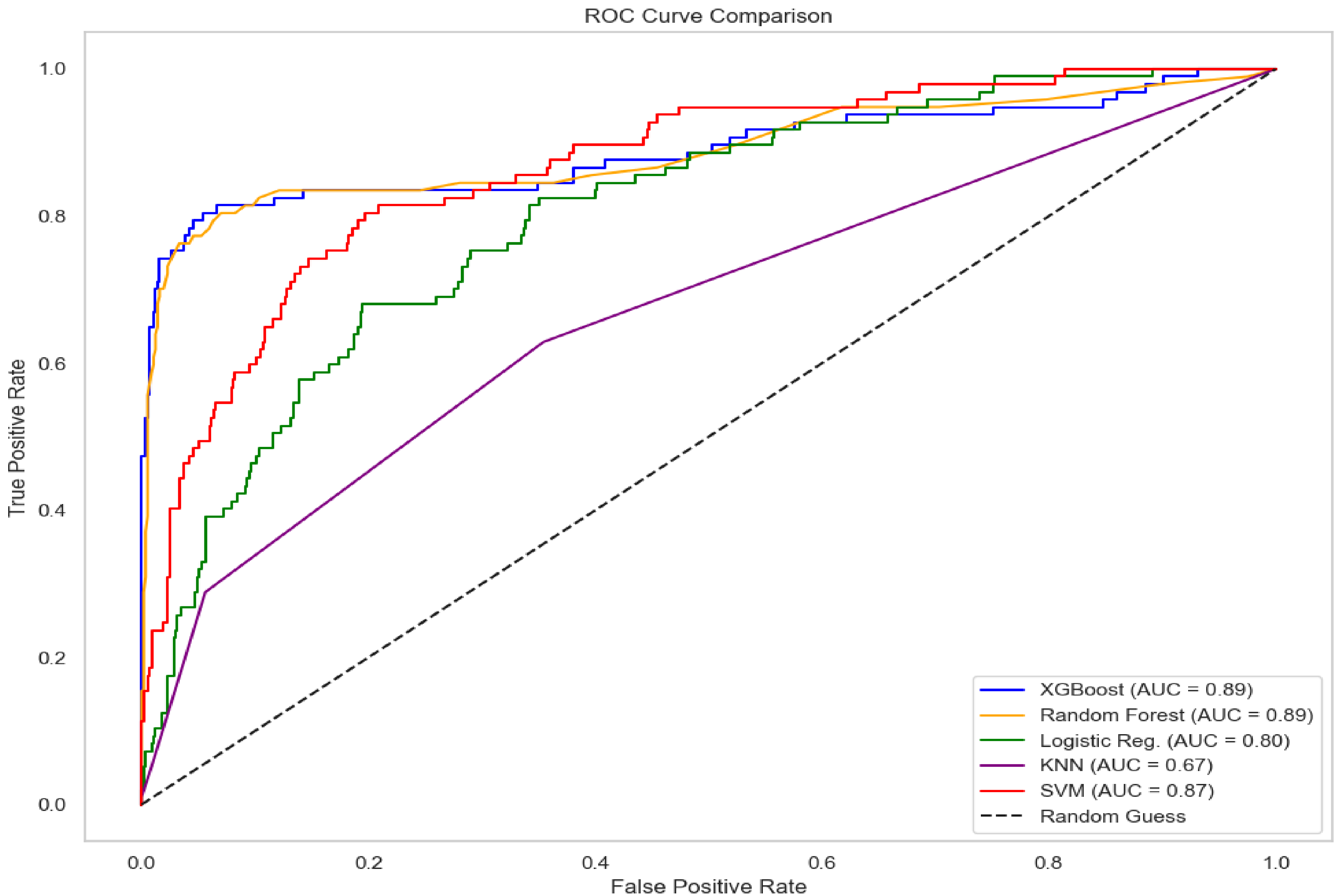
•**Recommendation**: Prioritize **XGBoost** for deployment due to its balanced and high performance across key evaluation metrics.

ROC Curve Comparison

AUC Comparison of Models

- XGBoost: 88.95%
- Random Forest: 88.92%
- SVM: 86.54%
- Logistic Reg.: 80.15%
- KNN: 67.09%

ROC Curve Image



Model Recommendation Based on ROC Analysis

Based on the ROC curves and AUC values, **XGBoost** and **m Forest** consistently demonstrate superior classification performance (**AUC = 0.89**).

•Their ability to accurately distinguish between classes makes them reliable for deployment in predictive tasks.

•It is therefore recommended to prioritize **XGBoost or Random Forest** for further development and integration.

•Models such as SVM and Logistic Regression offer reasonable performance but fall short of the top contenders.

•**KNN is not advised** due to its relatively low predictive capability (**AUC = 0.67**).

Best Model - XGBoost

XGBoost Performance:

- - Accuracy: 94.9%
- - Precision: 88.9%
- - Recall: 74.2%
- - AUC: 88.95%

Recommendations

- **Promote loyalty programs for month-to-month users**

- Incentivize long-term contracts over month-to-month plans to reduce churn

- **Encourage AutoPay Enrollment**

- Offer discounts or perks to customers who use AutoPay to improve retention

- **Enhance Fiber Optic Services**

- Address complaints and improve quality for Fiber Optic users—key to reducing dissatisfaction.

Next Steps

What's Next:

<u>Model Deployment:</u> Launch the best-performing churn prediction model (XGBoost) via an API.

<u>CRM Integration:</u> Connect the model to SyriaTel's CRM system for real-time churn alerts.

<u>Continuous Improvement:</u> Set up a feedback loop to **monitor**, **retrain**, and **update** the model regularly using fresh customer data.

Conclusion

- Our analysis of SyriaTel's customer churn data reveals that churn is influenced by contract type, payment method, and service quality—particularly for Fiber Optic customers. By leveraging machine learning models, we achieved high prediction accuracy, with **XGBoost performing best** (AUC: **88.95%**). These insights offer a data-driven foundation to proactively **reduce churn, enhance customer satisfaction**, and support strategic decision-making.

Thank You

Questions?