Predictive Analysis for customer churn in SyriaTel

BY MICHELLE NYAANGA

Overview

This project aims to address the inefficient management of customer churn within SyriaTel's subscriber base, where traditional churn prediction methods have proven suboptimal, leading to ineffective resource allocation and missed opportunities for retaining valuable customers. To solve this problem, the project leverages advanced analytics and machine learning techniques to develop a predictive model that accurately identifies customers at risk of churn.

Business Understanding

SyriaTel, a telecommunications company, is experiencing high customer churn rates, which means customers are stopping their service and going to competitors. This leads to lost revenue and potential decline in market share.

The company wants to reduce customer churn to increase their revenue and customer retention, which will be done through analysing historical customer data and using advanced analytics and predictive modelling.

By predicting which customers are likely to leave, SyriaTel can take proactive measures to retain them, strengthening its position in the telecommunications industry.

Problem Statement

SyriaTel,is experiencing high customer churn rates impacting revenue streams and market competitiveness despite considerable investments in marketing and retention strategies.

This project aims to address the inefficient management of customer churn within SyriaTel's subscriber base, where previous churn prediction methods have proven suboptimal, leading to ineffective resource allocation and missed opportunities for retaining valuable customers.

To solve this problem, the project leverages advanced analytics and machine learning techniques to develop a predictive model that accurately identifies customers at risk of churn

Objectives

Main

The primary goal of this project is to develop a churn prediction model that accurately predicts customers churn in SyriaTels's subscriber base by employing advanced analytics and machine learning techniques.

Specific Objectives

1. To Analyse the historical Data
2. To Develop a Predictive Model
3. To Implement Retention Strategies

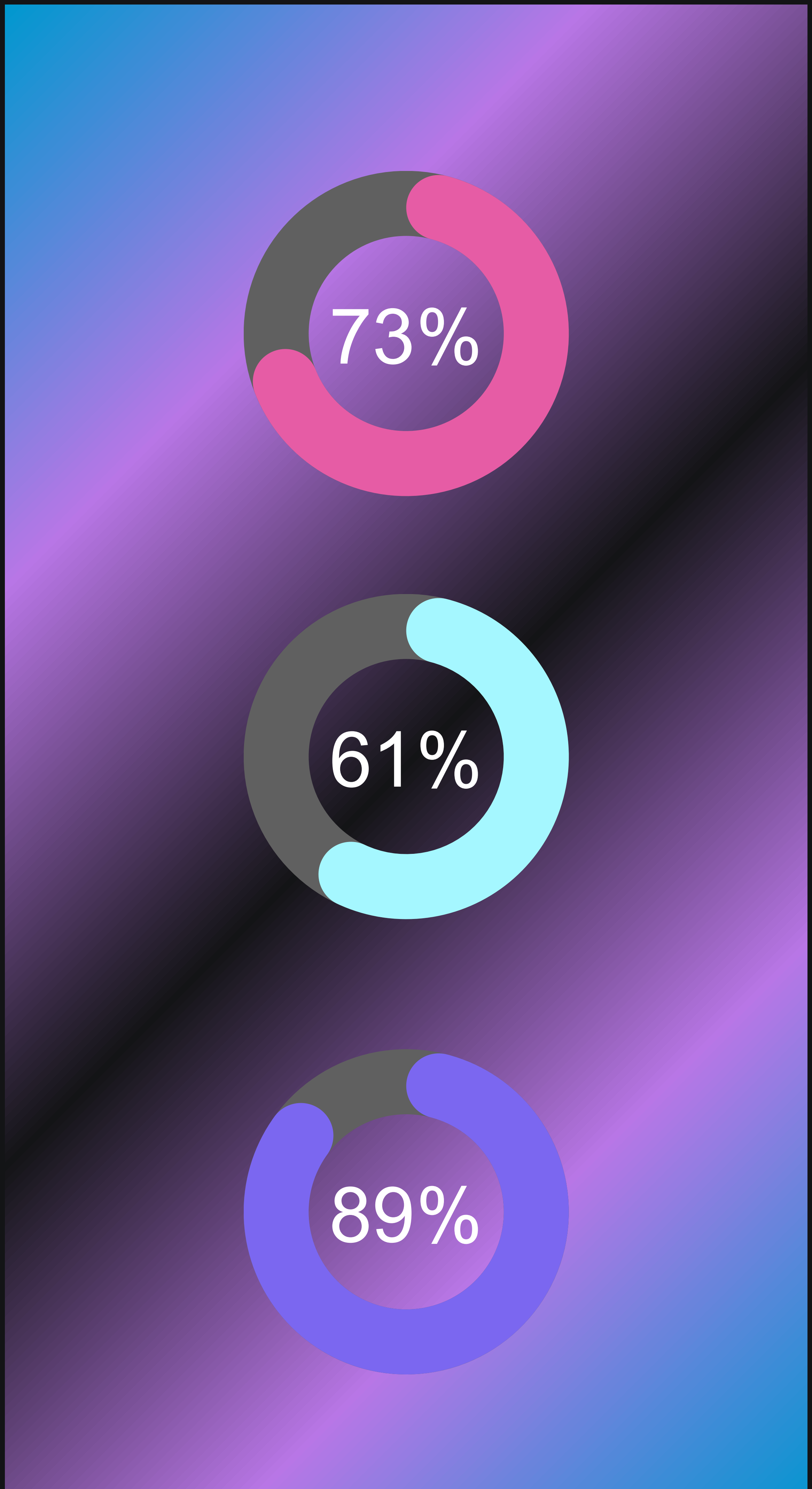
Data Understanding

The dataset is sourced from Kaggle.

It provides information on the customers behaviors which enables analysis and prediction of the churn patterns.

It contains 3333 entries and 21 features.

All features, except for Phone number and State, contain numerical values, while the remaining features are categorical or binary



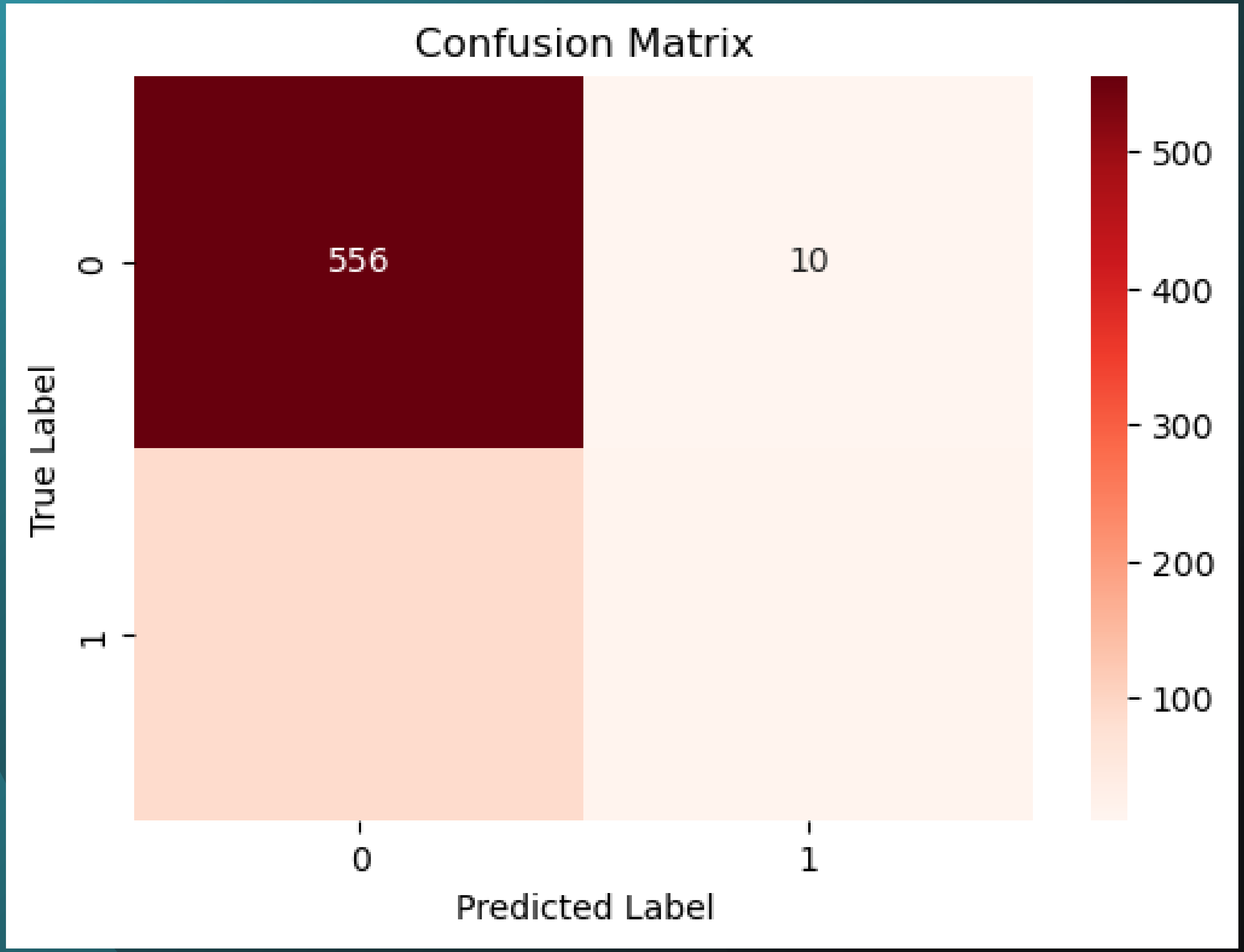
Data Preparation

Steps include:

- Data Cleaning
- Exploratory Data Analysis
- Data Processing

Modeling

01 Baseline model: Logistic Regression



The model achieves an overall accuracy of about 84.56%.

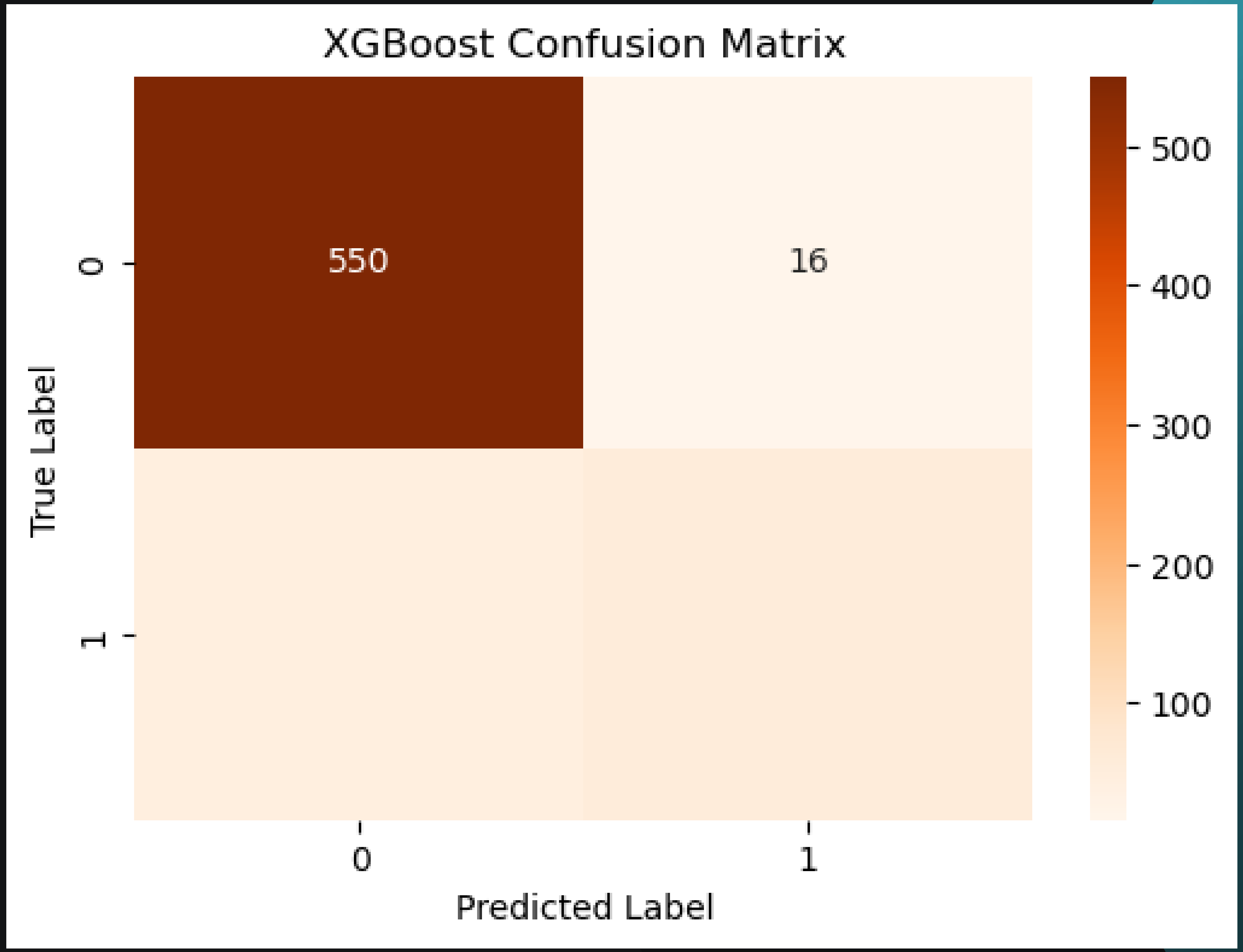
The precision is 0.46, meaning that 46% of the instances predicted as positive are actually positive.

The recall is 0.11, indicating that only 11% of the actual positive instances are correctly identified.

The F1-score balances precision and recall.

The model performs well in identifying true negatives (non-churners) but has difficulty predicting true positives (churners).

02 XGboost Classifier



The model achieves an overall accuracy of around 87.71%,

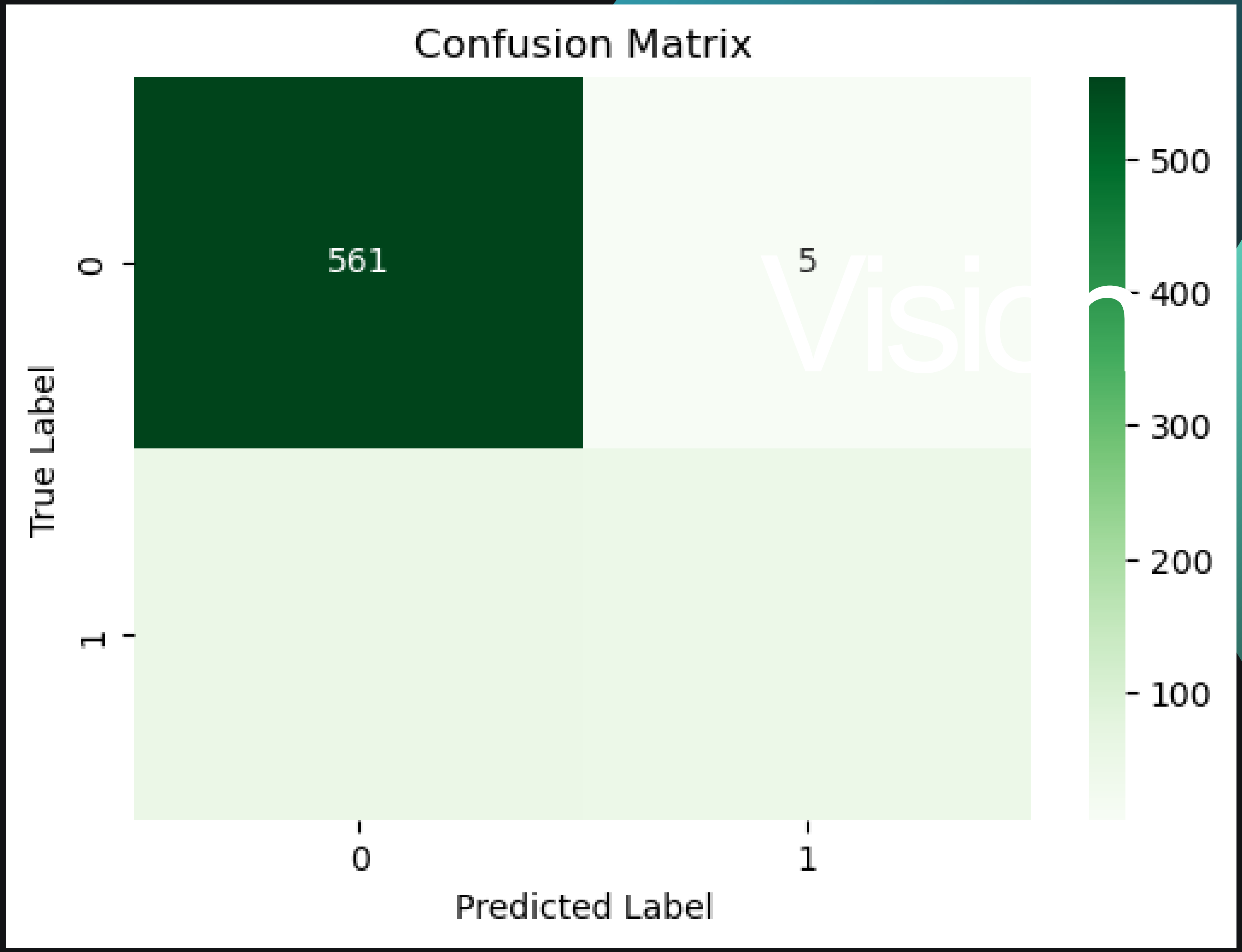
The precision is 0.69, meaning 69% of the instances predicted as positive are correct.

The recall is 0.34, indicating that 34% of the actual positive instances are accurately identified.

The F1-score is 0.45, reflecting moderate performance in predicting positive instances accurately.

This model performs well in identifying true negatives (non-churners) but has difficulty accurately predicting churners.

03 Gradient Boosting Model



The model has an overall accuracy of about 88.16%.

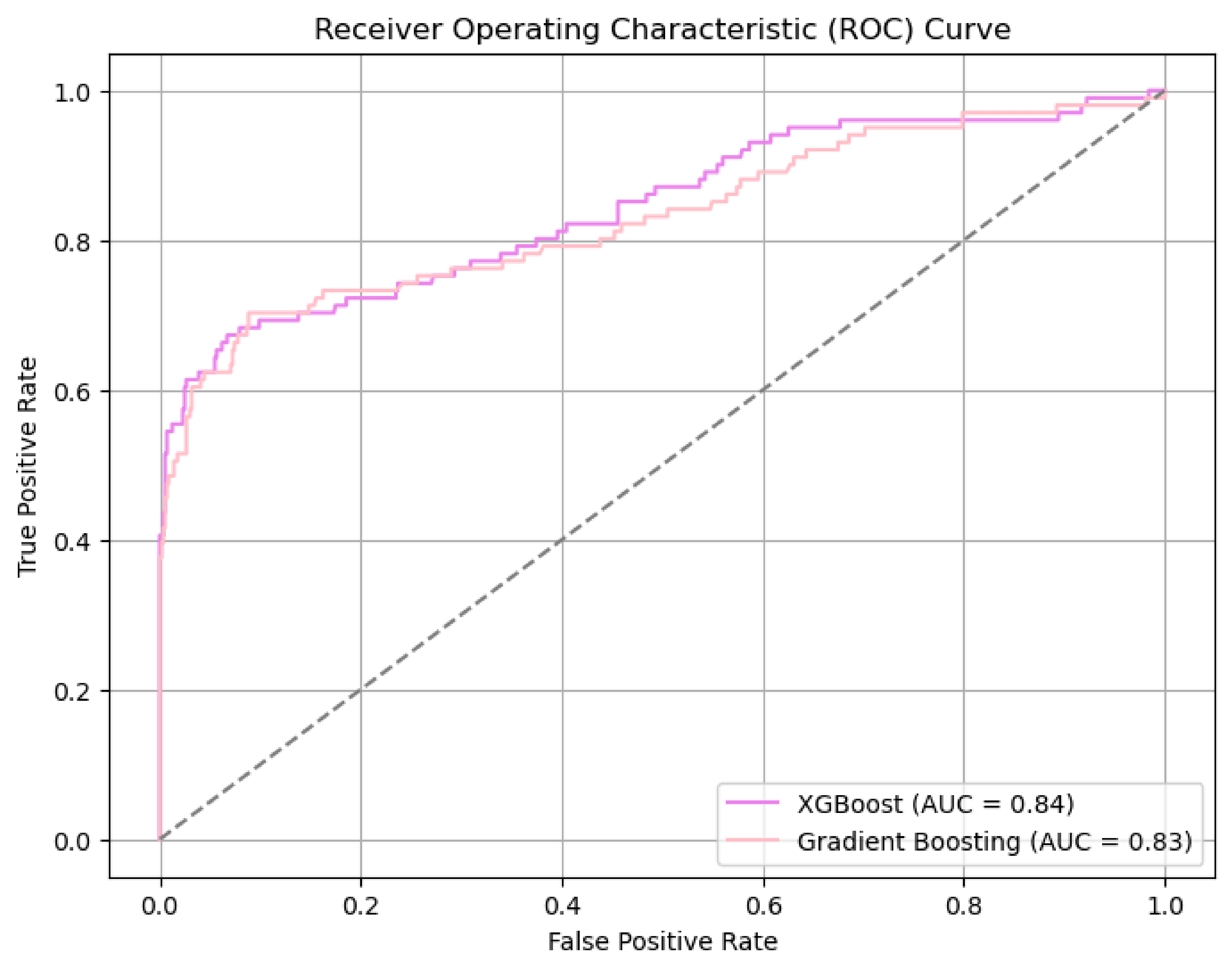
The precision is 0.78, showing that 78% of the instances predicted as positive are correct.

The recall is 0.31, indicating that 31% of the actual positive instances are accurately identified.

The F1-score is 0.44, indicating moderate performance in correctly predicting positive instances.

The model excels at predicting true negatives (non-churners) but struggles with accurately predicting churners.

Tuning the best two models



The optimal ROC curve in the graph corresponds to the Gradient Boosting model, indicating its superior performance by achieving the best balance between correctly identifying positive instances and minimizing false positives.

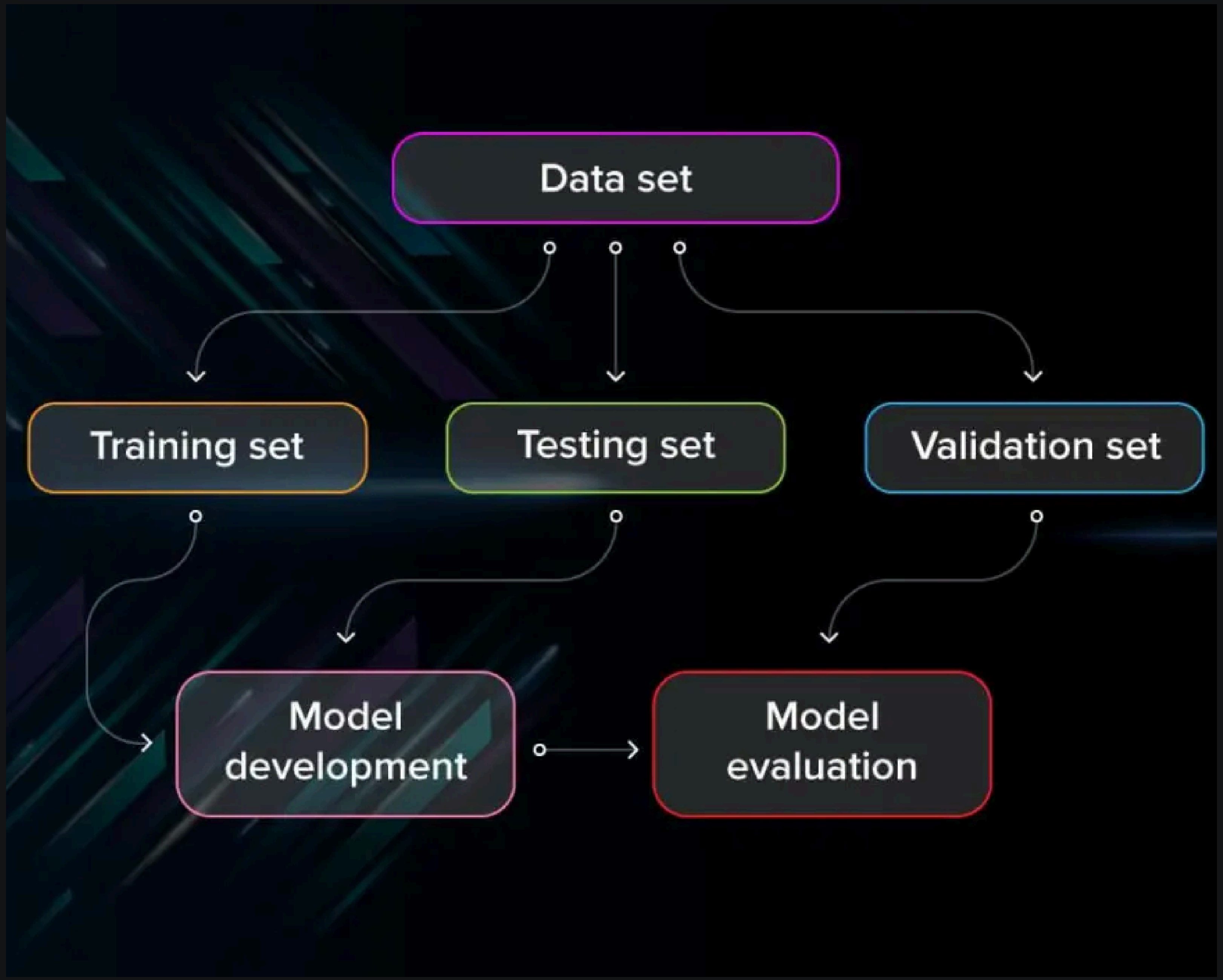
Evaluation

Three models were tested: Logistic Regression, Gradient Boosting, and XGBoost.

After evaluation, two models were fine-tuned for improved performance.

The Test ROC AUC Score measures the model's ability to distinguish between positive and negative outcomes.In this case, Gradient Boosting had the highest score of 0.76, indicating superior performance in differentiating between outcomes.

Gradient Boosting outperformed the other models, demonstrating higher accuracy and a better balance between true positives and false positives. It effectively identifies customers likely to leave while minimizing false positives.



Conclusion

Gradient Boosting outperformed the other models, demonstrating higher accuracy and a better balance between true positives and false positives

Several features such as the total day minutes, total night minutes, total eve minutes, international plan and voicemail plans are key predictors of churn.

Recommendations

- Introduce Loyalty Programs and Offers.
- Maintain Regular Communication.
- Customise Customer Experience.
- Predict and Prevent Churn.
- Collect Customer Feedback.

Next Steps

- Expand Data Collection
- Deploy the Model
- Monitor and Update the Model

Contact

www.reallygreatsite.com

michelle.nyaanga@student.moringaschool.com

0790110629

Nairobi West

Thank You