Sharon-Mukami / Phase_3_Project

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

MIT license

0 stars 0 forks 1 watching 1 Branch 0 Tags Activity

Public repository

1 Branch 0 Tags Go to file t Go to file + Add file Code

Sharon-Mukami Complete project c607384 · 1 minute ago

Data	Complete project	1 minute ago
.gitignore	Initial commit	18 hours ago
LICENSE	Initial commit	18 hours ago
Phase_3_project.ipynb	Complete project	1 minute ago
README.md	Update README.md	18 minutes ago
SYRIATEL PREDICITVE ANALY...	add presentation pdf	14 minutes ago
image 1.png	add files	17 hours ago
image 2.png	add files	17 hours ago
image 3.png	add files	17 hours ago
image 4.png	add files	17 hours ago
image.jfif	add files	17 hours ago

README MIT license

Phase_3_Project

Predictive Analysis for customer churn in Syria Tel.

You can access the presentation slides here

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 (meaning 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 telecommunication industry.

The key stakeholders include: SyriaTel Management whose interests are strategies to reduce churn. Marketing Team whose interests are holding campaigns for customer retention towards the atrisk customers.

Problem Statement

Syria Tel wants to predict customer churn based on historical data to identify customers at risk of leaving the service. By doing so, the company can implement targeted retention strategies to reduce customer attrition and improve overall customer satisfaction.

Objectives

1. Analyse the historical Data: Performing a thorough analysis of SyriaTel's historical customer data, which include usage patterns, service interactions, and churn records, to identify key features and trends associated with potential churn.

2. Develop a Predictive Model: Using advanced analytics and machine learning algorithms, such as logistic regression and ensemble methods, to construct a predictive model with high accuracy for forecasting customer churn.

3. Develop and Implement Retention Strategies: Integrating the predictive model into SyriaTel's operational framework to facilitate real-time detection of at-risk customers. Developing and also implementing personalised retention strategies based on the model's predictions to effectively reduce churn.

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.

Summary of the datasets features:

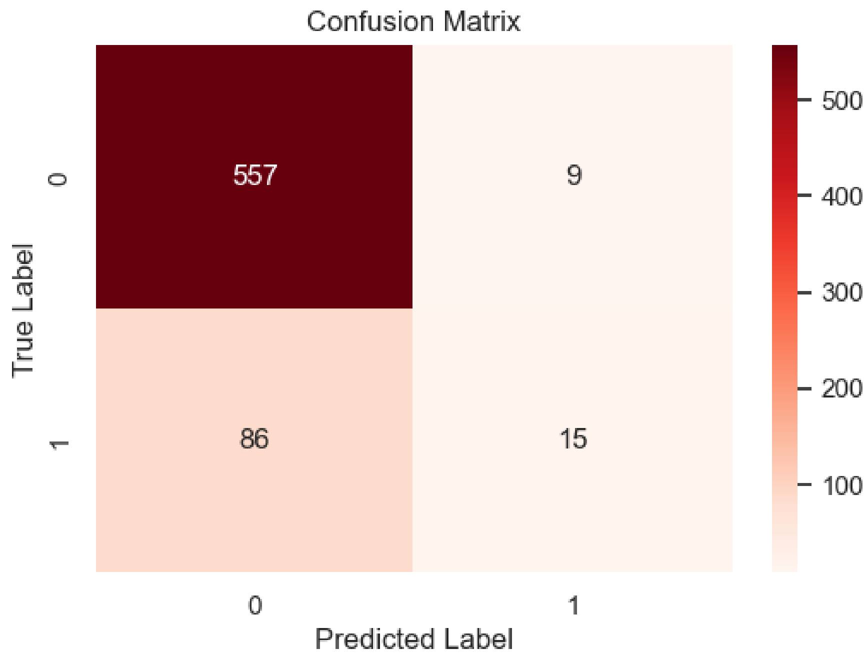
- State: The state in which the customer resides
- Account length: The number of days the customer has been the company.
- Area code: The area code of the customer's phone number.
- Phone number: The customer's phone number.
- International plan: Whether the customer has an international plan or not(Y/N)
- Voice mail plan: Whether the customer has a voicemail plan or not(Y/N)
- Number vmail messages: The number of voicemail messages.
- Total day minutes: Total number of minutes the customer used during the day.
- Total day calls: Total number of calls the customer made during the day.
- Total day charge: Total charges for calls made during the day.
- Total eve minutes: Total number of minutes the customer used during the evening.
- Total eve calls: Total number of calls the customer made during the evening.
- Total eve charge: Total charges for calls made during the evening.
- Total night minutes: Total number of minutes the customer used during the night.
- Total night calls: Total number of calls the customer made during the night.
- Total night charge: Total charges for calls made during the night.
- Total intl minutes: Total number of international calls made.
- Total intl calls: Total number of international calls made.
- Total intl charge: Total charges for international calls.
- Customer service calls: Number of customer service calls made.
- Churn: Whether the customer churned or not(Y/N).

Modeling

1. Baseline model: Logistic Regression: Accuracy: The model achieves an overall accuracy of about 84.56%, correctly classifying approximately 84.56% of the test dataset instances.

Precision and Recall: For class 1, the precision is 0.46, meaning that 46% of the instances predicted as positive are actually positive. The recall for class 1 is 0.11, indicating that only 11% of the actual positive instances are correctly identified.

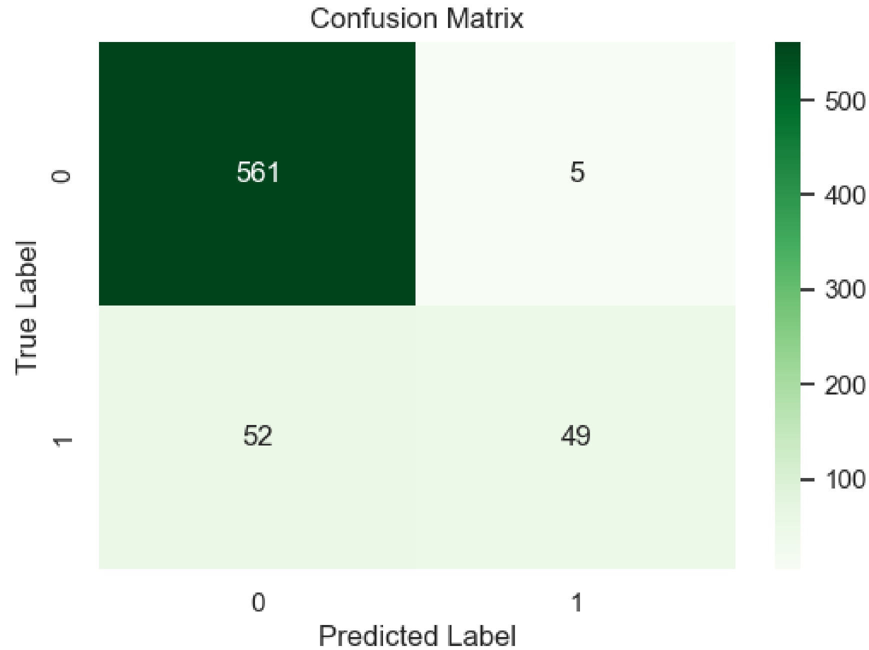
F1-score: The F1-score balances precision and recall. For class 1, the F1-score is 0.18, highlighting the model's poor performance in accurately predicting positive instances. The model performs well in identifying true negatives (non-churners) but has difficulty predicting true positives (churners).



2. Gradient Boosting Model: Accuracy:The model has an overall accuracy of about 88.16%, meaning it correctly predicts the class label for 88.16% of the test dataset instances.

Precision and Recall: For class 1, the precision is 0.78, showing that 78% of the instances predicted as positive are correct. The recall for class 1 is 0.31, indicating that 31% of the actual positive instances are accurately identified.

F1-score: The F1-score for class 1 is 0.44, indicating moderate performance in correctly predicting positive instances.The model excels at predicting true negatives (non-churners) but struggles with accurately

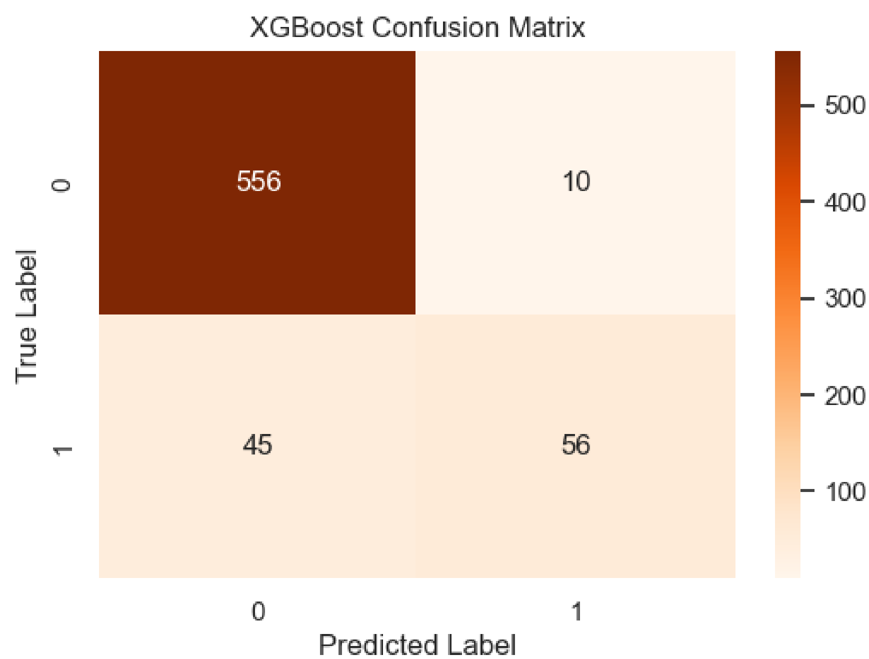


predicting churners.

3. XGboost Classifier: Accuracy: The model achieves an overall accuracy of around 87.71%, accurately predicting the class label for 87.71% of the test dataset instances.

Precision and Recall: For class 1, the precision is 0.69, meaning 69% of the instances predicted as positive are correct. The recall for class 1 is 0.34, indicating that 34% of the actual positive instances are accurately identified.

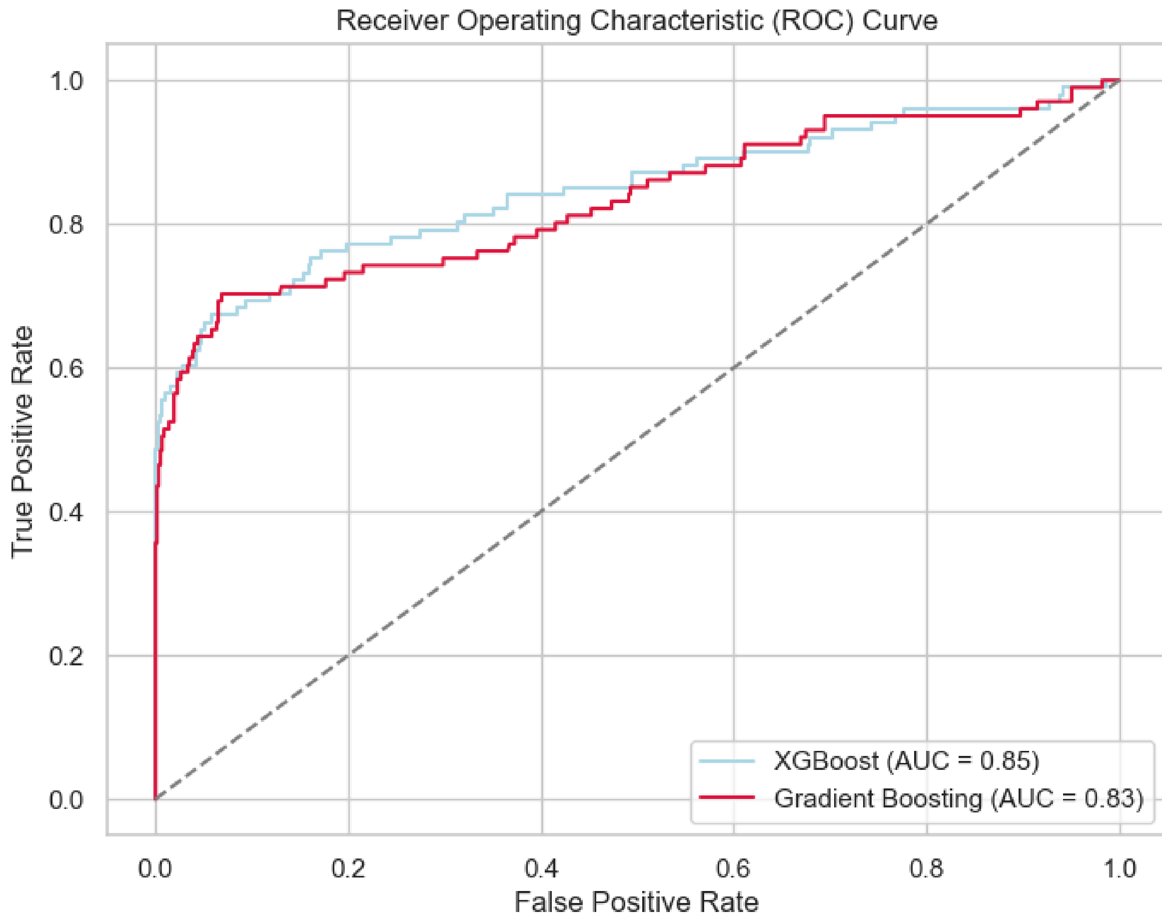
F1-score: The F1-score for class 1 is 0.45, reflecting moderate performance in predicting positive instances accurately. This model performs well in identifying true negatives (non-churners) but has difficulty accurately



predicting churners.

4. 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. It effectively identifies customers likely to leave while minimizing 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. Higher call durations during the day, night and evening increase the likelihood of churn. Customers without an international plan and a voice mail plan are more likely to churn.

Recommendation

- Introduce Loyalty Programs and Offers: Implement loyalty programs, exclusive offers, and perks to incentivize customer retention. Provide discounts, free upgrades, or access to premium content to reward long-term customers.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%