

Predicting Customer Retention at Syria-Tel Mobile Telecom using Machine Learning

Presented by: DAISY WAGATI

Project Overview

In an increasingly competitive telecommunications landscape, Syria-Tel is taking proactive steps to combat customer churn and enhance retention. This project focuses on developing a robust binary classification model that predicts customer churn by uncovering underlying patterns in customer behavior.

By leveraging data-driven insights, Syria-Tel aims to implement targeted retention strategies that not only mitigate revenue loss from attrition but also foster deeper customer loyalty.

Business Understanding

In the fast-paced telecommunications industry, customer retention is crucial for maintaining profitability and competitive advantage. Syria-Tel faces the pressing challenge of customer churn, which not only impacts revenue but also threatens brand loyalty and market share.

This project aims to bridge the gap between data analytics and business objectives by identifying key drivers of churn through predictive modeling. By gaining insights into customer behavior and preferences, Syria-Tel can tailor its services and communication strategies to meet the needs of its customers, ultimately enhancing satisfaction and loyalty.

Research Questions

- Create and validate a predictive model to accurately identify customers who are likely to churn in the near future, using historical customer data and relevant features.
- Analyze customer data to identify the key factors and behaviors that contribute to the likelihood of churn.
- Design and implement targeted interventions for identified at-risk customers, and assess the effectiveness of these interventions in improving customer satisfaction and reducing churn rates.

Data Understanding

- The telecom dataset contains information about customer activity in relation to churn.
- The dataset contains 3333 entries and 21 columns.
- It contains both numerical and categorical columns.
- The data is a binary classification model.
- There are no missing values or duplicates.

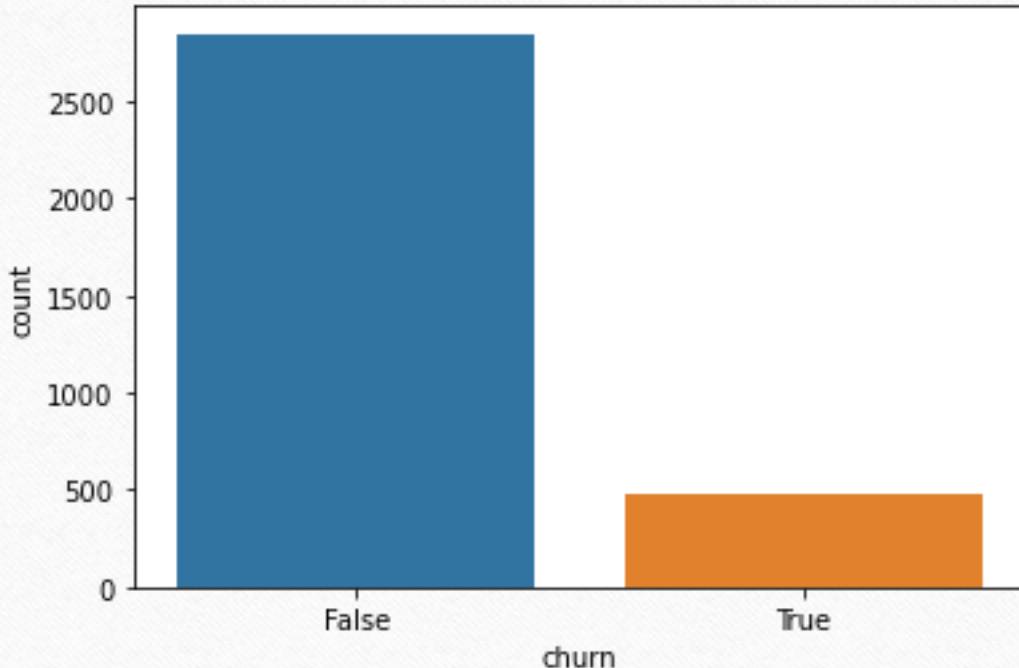
Exploratory Data Analysis

Of the 3,333 customers in the dataset, 483 have terminated their contract with SyriaTel.

That is 14.5% of customers lost.

True = Churned

False = Not Churned



Modelling

Three different models were built to predict churn and further improvements were made through hyper-parameter tuning to achieve the best predictive results.

The models used are:

- Logistic Regression Model
- Decision Tree Classifier
- Random Forest Classifier

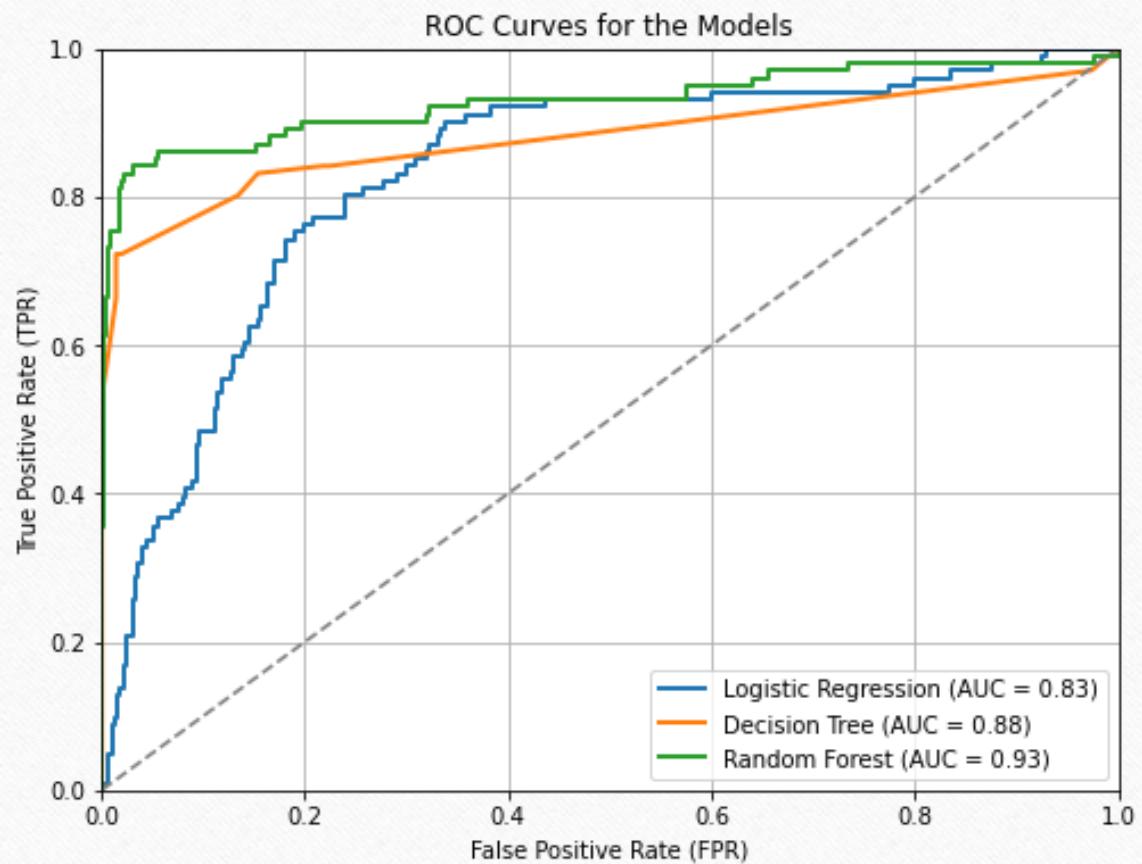
Comparison Frame

Upon analysis, we found that the Random Forest Classifier was the best model to use. It has the highest F1 score and precision.

	Model	Accuracy (Test Set)	F1 Score (Test Set)	Recall (Test Set)	Precision (Test Set)
0	Logistic Regression	0.780000	0.510000	0.770000	0.390000
1	Decision Trees Classifier	0.950000	0.802000	0.720000	0.900000
2	Random Forest Classifier	0.940000	0.804000	0.690000	0.960000

ROC and AUC curves

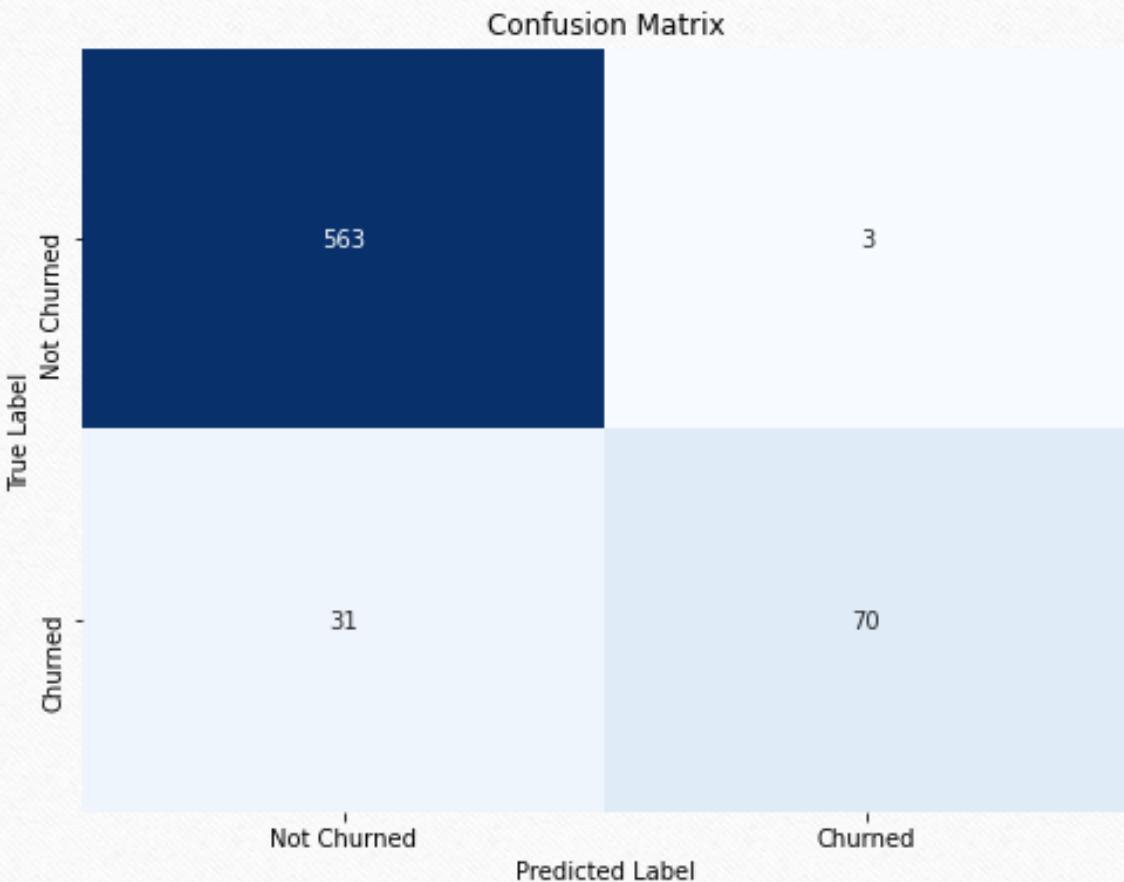
Upon further analysis, the Random Forest model outperformed the others, showing a higher Area Under the Curve (AUC) and better classification performance, making it the most effective model for the given task.



Model Selected: Random Forest Classifier with reduced estimators

The model achieved:

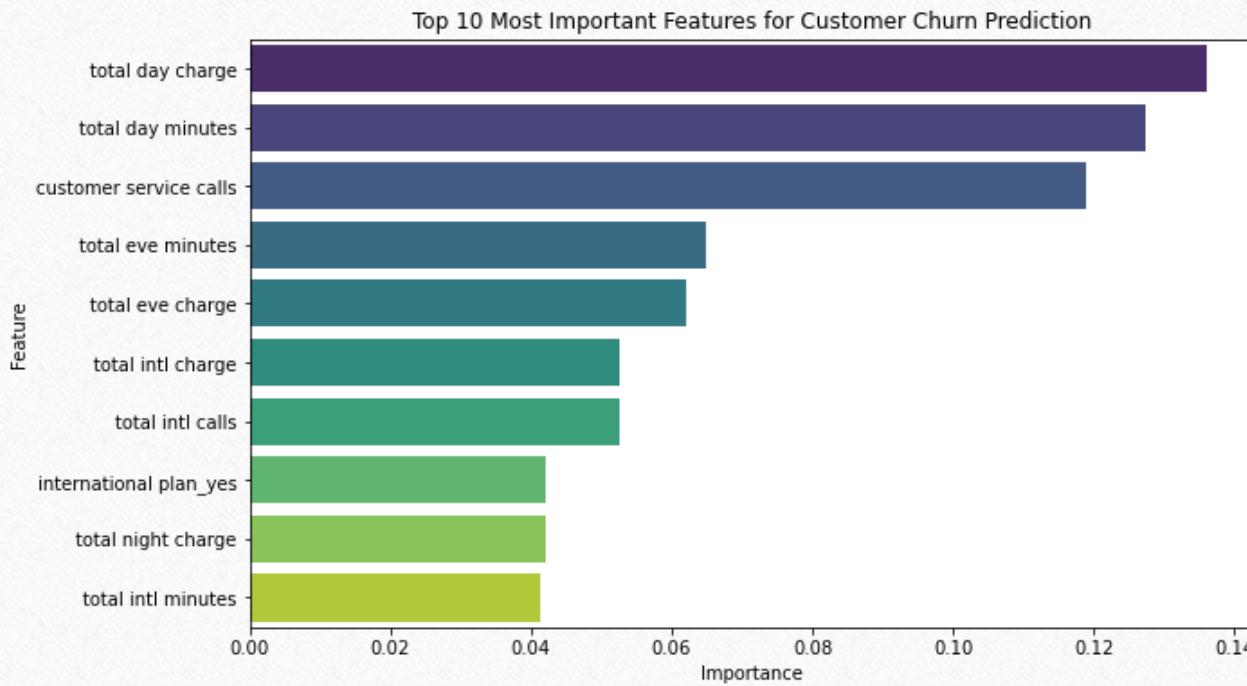
- Accuracy: 94.37%
- Precision: 95.89%
- Recall: 69.3%
- F1 Score: 80.45%
- Train score: 97.71%
- Test Score: 94.9%



Model Interpretation

- The Model's accuracy, reflecting its ability to reflect churn, is 94.37%
- Precision measures the accuracy of positive predictions, and a precision of 95.89% indicates that when the model predicts a positive outcome, it is correct 95.89% of the time.
- Recall represents the model's ability to predict actual positive instances , this indicates that the model captures 69.3% of the total positive instances.
- The F1 score, which combines precision and recall , is 80.46%, indicating a balanced performance between precision and recall for the positive predictions made by the model.
- Given the train score of 97.7% and Test score of 94.9%, the model does not overfit, demonstrating its ability to generalize and make accurate predictions on new and unseen data.

Feature Importance



Recommendations

- **Adopt Random Forest Classifier:** Utilize the Random Forest Classifier as the primary model for predicting customer churn, leveraging its superior performance metrics such as ROC curve, accuracy, F1-score, recall, and precision to effectively identify at-risk customers.
- **Enhance Customer Service Training:** Invest in training customer service representatives to address customer concerns with greater efficiency and empathy. Implement customer feedback mechanisms to proactively identify and resolve pain points, reducing the likelihood of churn. Additionally, explore the use of customer feedback mechanisms to proactively identify and resolve pain points before they lead to churn.
- **Revise Pricing Structure:** Review and adjust pricing for total day charges to ensure competitiveness and transparency. Introduce flexible plans that incentivize evening usage, such as discounts or evening minutes, to enhance customer satisfaction and encourage off-peak usage, ultimately improving retention rates.



Thank You

Any Questions?

Git Hub: <https://github.com/Daisy96494/Phase3ProjectDaisy>