Predicting Customer Churn for SyriaTel: Findings and Recommendations

Overview

The overall goal of the project is to develop a binary classification model that will help SyriaTel determine whether a customer will abandon their services based on information provided in the dataset.

Business Understanding

Customer churn is a significant challenge in the telecommunications industry. Addressing churn is crucial as acquiring customers is more expensive than retaining existing one.

SyriaTel is currently grappling with challenges related to customer retention. They seek to understand and predict customer churn to enhance its competitive advantage and identify which customers are at risk of leaving.

Business Questions

- What are the key factors leading to customer churn for SyriaTel?
- What is the best predictive model that can be used to predict churn based on the available data?
- What strategies can SyriaTel implement to reduce churn and retain customers?
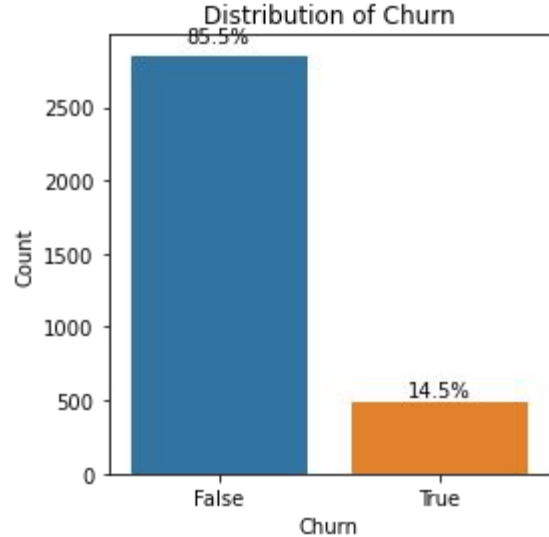
Data Understanding

- The dataset contains information about customer activity in relation to churn.
- The dataset contains 3333 rows and 21 columns with the churn column being the column of interest indicating whether the customer churned or not.
- The dataset contains both numerical and categorical values.
- The dataset contains no missing values or duplicate values.
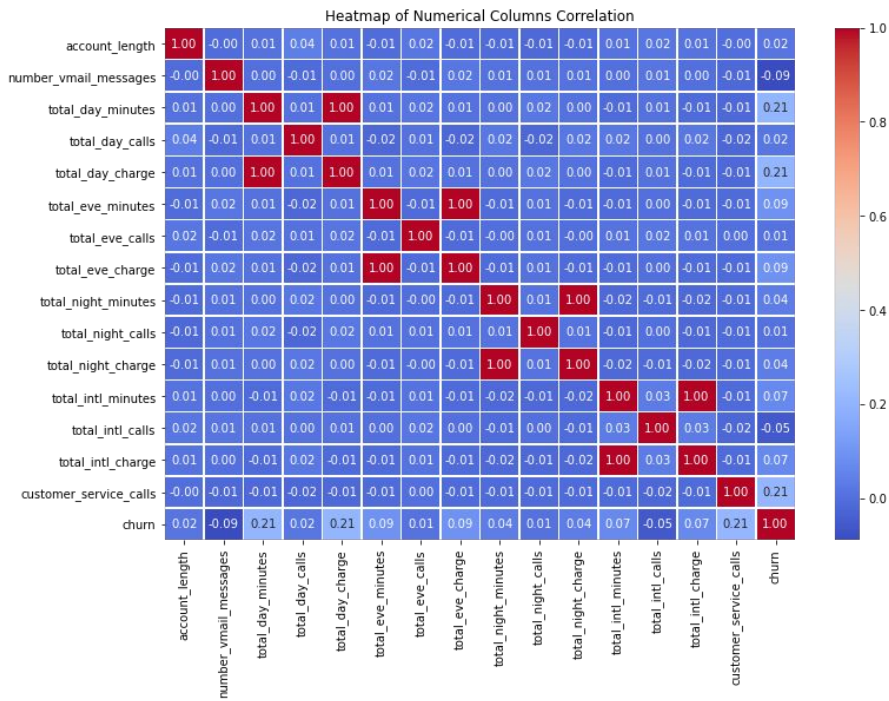- The problem is a binary classification problem.

Data Understanding

Of the 3333 customers in the dataset, 483 customers abandoned the services offered by SyriaTel. Here:

- False = not churned
- True = churned

Data Understanding

Heatmap of Numerical Columns Correlation

Features that are positively correlated with churn:

- Total day charge
- Total day minutes
- Customer service calls
- Total international charge
- Total evening charge
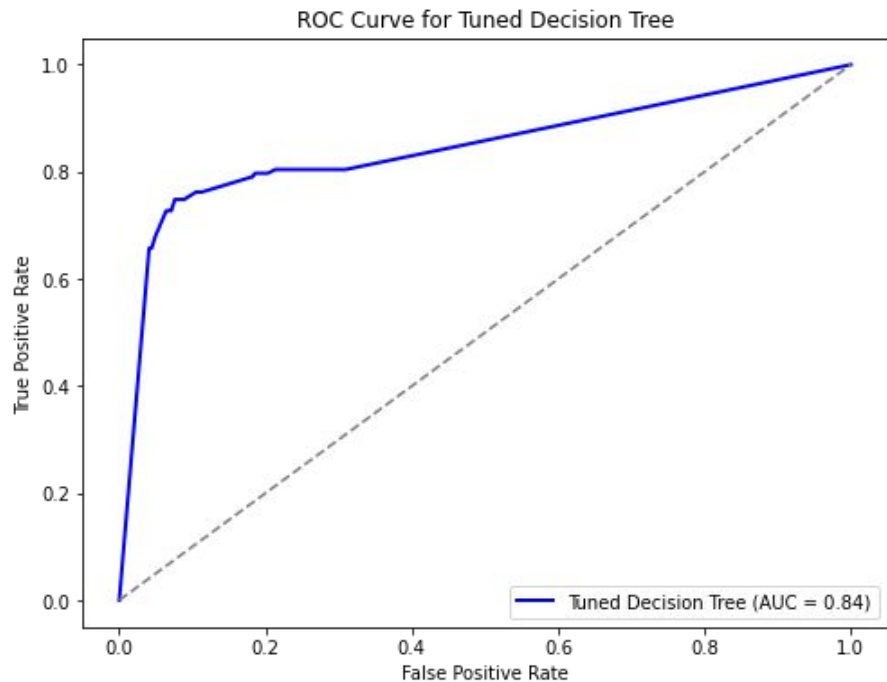- Total evening minutes

Modelling

Two models were built to predict churn and iterated upon to achieve the best result. The models built were:

- Logistic Regression Model
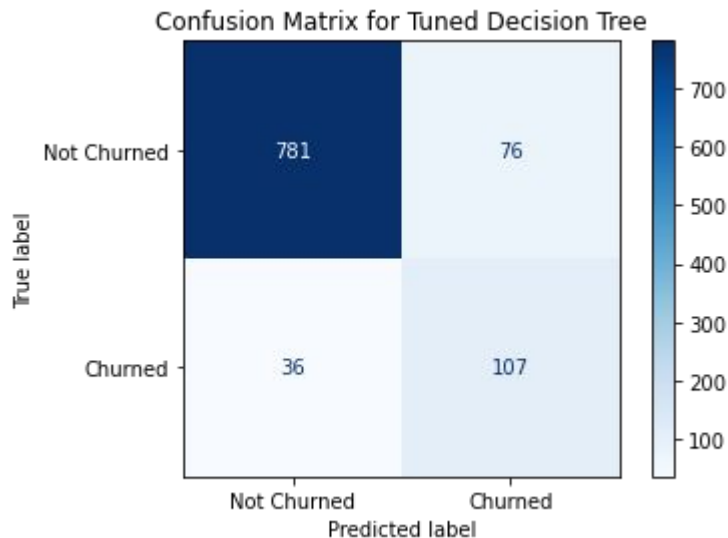- Decision Tree Classifier

Evaluation

ROC Curve for Tuned Decision Tree

The best model is the tuned decision tree model with an accuracy of 89%.

ROC-AUC score of 84% shows the model has a good ability to distinguish between the two classes.

Evaluation

Confusion Matrix for Tuned Decision Tree

The model achieved:

- Accuracy: 89%
- Precision (not churned): 96%
- Precision (churned): 58%
- Recall (not churned): 91%
- Recall (churned): 75%
- F1-Score: 93%
- Test accuracy: 89%

Recommendations

- SyriaTel should adopt the Tuned Decision Tree model as it is preferable due to its ROC-AUC score as well as its ability to generalize well with unseen data i.e, test set.
- Focus on customer calls, total day charge, subscription to an international plan, total international calls and total evening charge when predicting if a customer will leave.
- Implement loyalty programs for long-tenure customers as they are less likely to leave.

Recommendations contd.

- Consider training customer service representatives to resolve issues more effectively on the first contact when possible.
- Offer flexible pricing options for day and evening packages since customers charged more tend to churn.

Next Steps

- Investigate additional features that can improve model performance and lead to customer churn.
- Consider other models to improve prediction performance.

Thank You!

Any Questions?