PROJECT
SYRIATEL CHURN

A MACHINE LEARNING CLASSIFICATION METRIC MODEL TO EFFECTIVELY PREDICT CUSTOMER CHURN

WHAT ARE FEATURES THAT LEAD TO CHURN?

CONTENTS

- OVERVIEW
- BUSINESS & DATA UNDERSTANDING
- MODELLING
- EVALUTION
- RECOMMENDATIONS

OVERVIEW

It is imperative to want to retain acquired customers as acquiring new customers involves huge marketing costs, that include huge advertising budgets and commissions to sales agents.

This project delves in the telecommunication industry with sourced data from Kaggle for SyriaTel using Classification prediction modelling, to predict the likelihood of loosing customer's business "Churn".

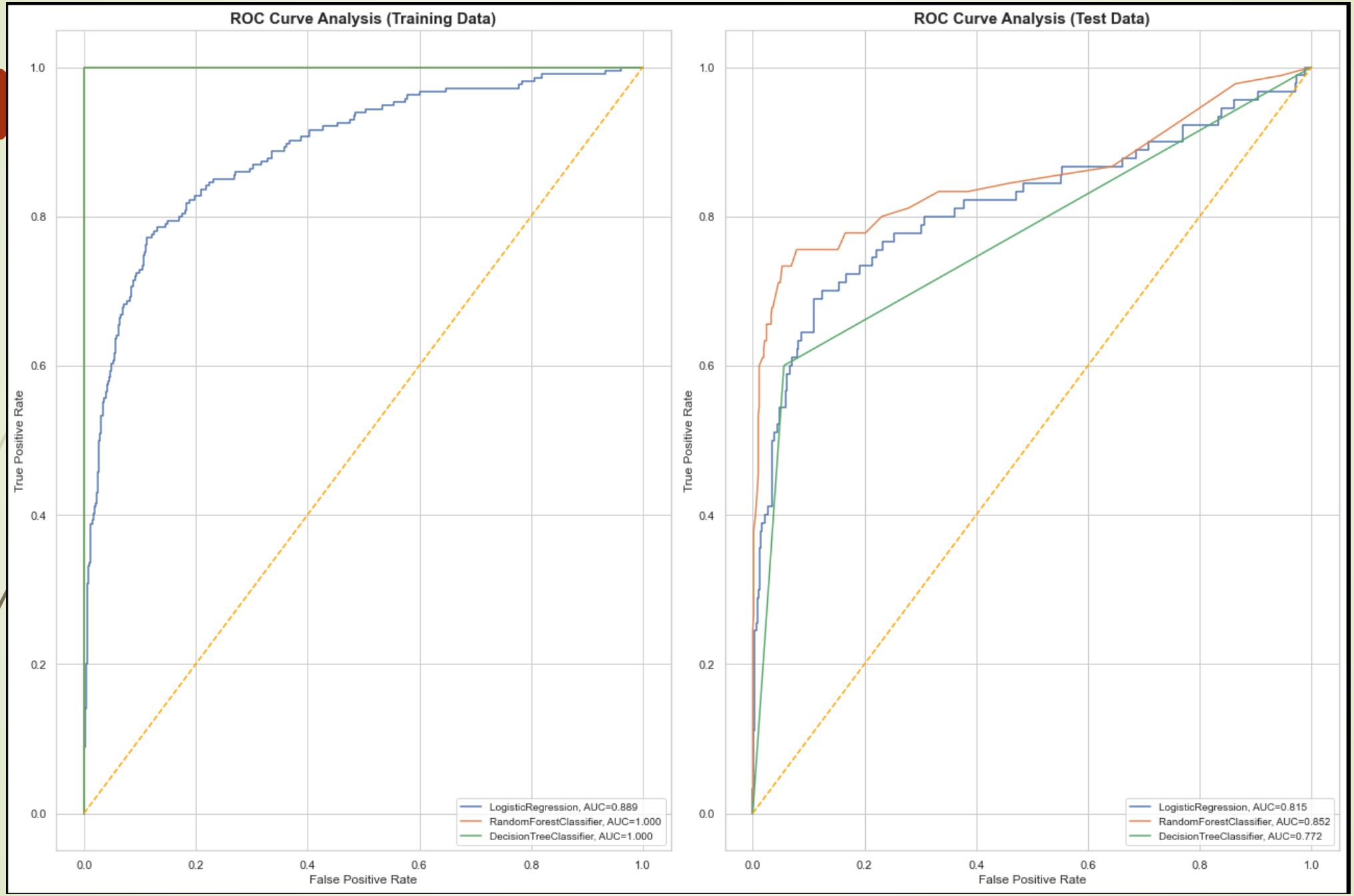
The company aims to go ahead of any hinderance that there can be and develop initiatives to ensure they retain the customer.

BUSINESS & DATA UNDERSTANDING

Lucky for us the data provided for Churn data is straight forward data that require minimal to zero iterations, containing various information about the customer.

With this presentation the company should be able to:

- Curate customer centric initiatives that will ensure retention of the customers.
- Identify problems that the customers meet and address.
- Curate different experientials for the customer to experience meeting their needs hence customer satisfaction.


ROC Curve Analysis (Training Data)
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
LogisticRegression, AUC=0.889
RandomForestClassifier, AUC=1.000
DecisionTreeClassifier, AUC=1.000
ROC Curve Analysis (Test Data)
True Positive Rate
False Positive Rate
LogisticRegression, AUC=0.815
RandomForestClassifier, AUC=0.852
DecisionTreeClassifier, AUC=0.772

- Accuracy is a measure of how often the model gets the prediction right, and in this case measures how often the model correctly predicts whether a customer will churn or not. An test accuracy score of 0.91 from the best Random Forest Ensemble Model means that our model was able to predict correctly 91% of the time.
- Random Forest and Decision Tree models have perfect accuracy and AUC scores of 1.000, suggesting they fit the training data perfectly. However, this could indicate overfitting, as these models may have memorized the training data rather than generalizing well.
- When evaluated on the test data, the Random Forest classifier stands out with the highest AUC (0.863270) and accuracy (92%). It outperforms the other two models, indicating better generalization and performance on unseen data. The Decision Tree model has a relatively high accuracy but a lower AUC, suggesting it may not handle the complexity of the data as well as Random Forest. Logistic Regression also has a lower AUC and accuracy compared to Random Forest.

Conclusion ;The Random Forest classifier is the best model to use, as it achieves the highest accuracy and AUC on both training and test data, with the best generalization capability to unseen data.

RECOMMENDATIONS

- ✓ The performance of the 3 models is indicative that the features in the SyriaTel data set have good predictive power.
- ✓ Provide targeted offers and discounts to customers based on their patterns and preferences enhancing retention.
- ✓ Analyze customer service interaction to identify common ailing issues to the customers and address each promptly.
- ✓ Address any issue that may hinder experience enhancement.
- ✓ Customer Service Calls also have an impact on churn, meaning a customer calling customer service more is more likely to churn.

Spencer Lugalia

Phase 3 Machine Learning Project