Predicting Customer Churn For SyriaTel Company

MASINDE VICTOR

Project Overview

- This project deals with a company, SyriaTel, that wants to know more about their customer churn.
- Customers leave a company due to different reasons and my project aims to uncover the reasons and predict customer churn.
- The company can then use the information gained from this project to work on retaining their customers.
- Retaining customers is cheaper and easier as compared to gaining a new customer.
- The project utilizes the SyriaTel Telecoms dataset to create a classification model that predicts if a customer will churn or not. This is a problem of binary classification.

Business understanding

- Communication is key in our daily lives as it is what keeps the world running.
- We have several comminucation network providers that help make it easy for us to interact with people.
- Competition gets tight as different companies try to intergrate the latest models into their systems for better customer satistfaction and better Return on Investments(ROI) on their end.
- One effect of these improvements is churning(customers leaving) as they propagate towards where they can get better services.
- I aim at creating a predictive business model which would enable Syriatel detect churn and identify reasons for the same.

Stakeholders

1. The Syriatel Telecommunications board.
2. The customers of the company.
3. The employees of the company.

Research objectives

- To identify the key features that determine if a customer is likely to churn.
- To determine the most suitable model to predict Customer Churn.
- To establish customer retention strategy to reduce churn.

Data understanding

- This dataset was sourced from kaggle and it has 3333 rows and 21 columns.
- The dataset has data recorded in different data type including float, intergers and objects.
- The columns are properly named showing what happens in the communication sector.
- This dataset includes details about a telecom company's customers, such as their state, phone number, area code, account length, and whether or not they have voice mail and international plans.
- Out of the 21 columns, 4 of them are in object form which means that they are categorical columns.

Modelling
1. Baseline Logistic Model

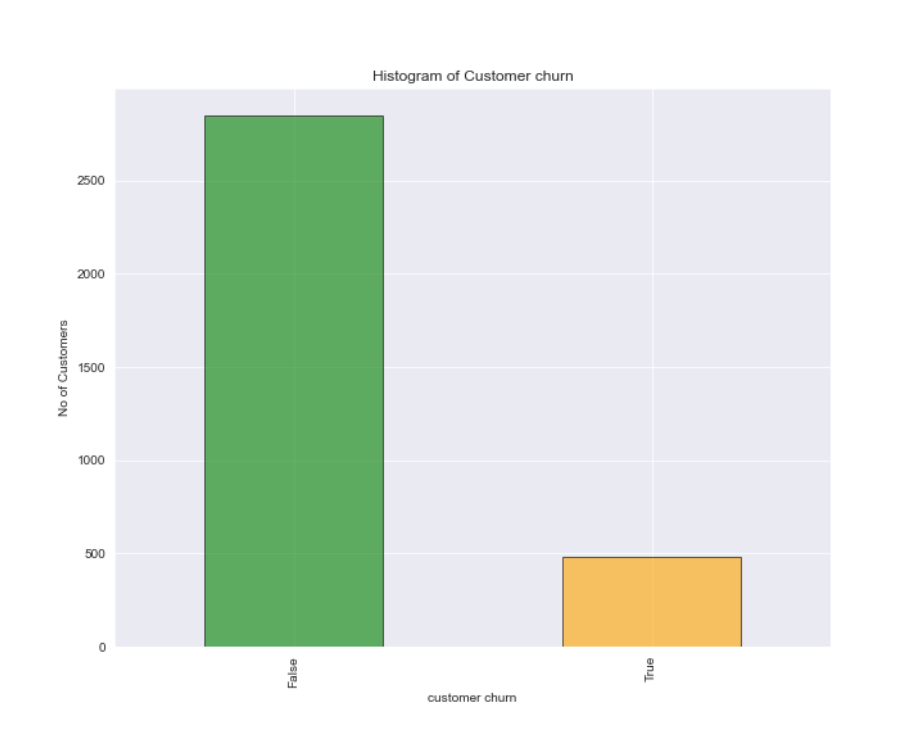
- Logistic regression is among the models that can be used when dealing with classiication problems.
- It requires that our target variable be in form of classes.
- For our case, we have a target variable with two classes commonly referred to as Binary class.
- This model if fitted with the data before any manipulation apart from splitting the data for train and test is done.
- We fit the model to scaled data of target(y) and feature variables(X).
- We can make predictions using the model and compare it with actual values to get the performance metrics of the model.

2. Tuned Logistic Model

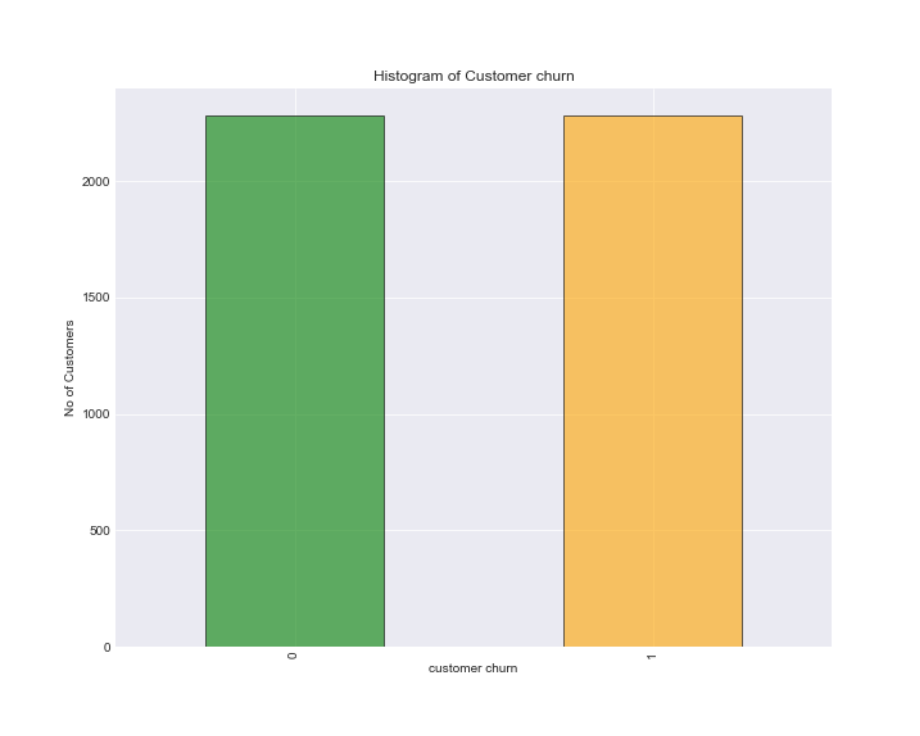
- There is an imbalance in our target class.
- The class `not churned`(0) is bigger than the `churned` class(1).
- This can lead to overfitting of the models and as a result predict one class very well but fail to predict the other.
- To address this we introduce SMOTE to help us in balancing the target classes.
- It does that by duplicating the minority class until they are even with the majority class.
- We then pass the balanced X and ys to our logistic model and make predictions which will in turn help us calculate the performance metrics of the model.

Shape of the Target Class

Before SMOTE



After SMOTE



3.Decision Tree Classifier

- Decision tree classifiers are sophisticated machine learning models that use tree-like structures to categorise data.
- They work by recursively splitting the feature space into smaller subsets based on the most informative attributes.
- Decision trees can handle category and numerical variables, making them useful tools for a variety of classification applications.
- For this model, we will make use of the data that have their classes balanced by SMOTE.
- We can also calculate the metrics of the model.

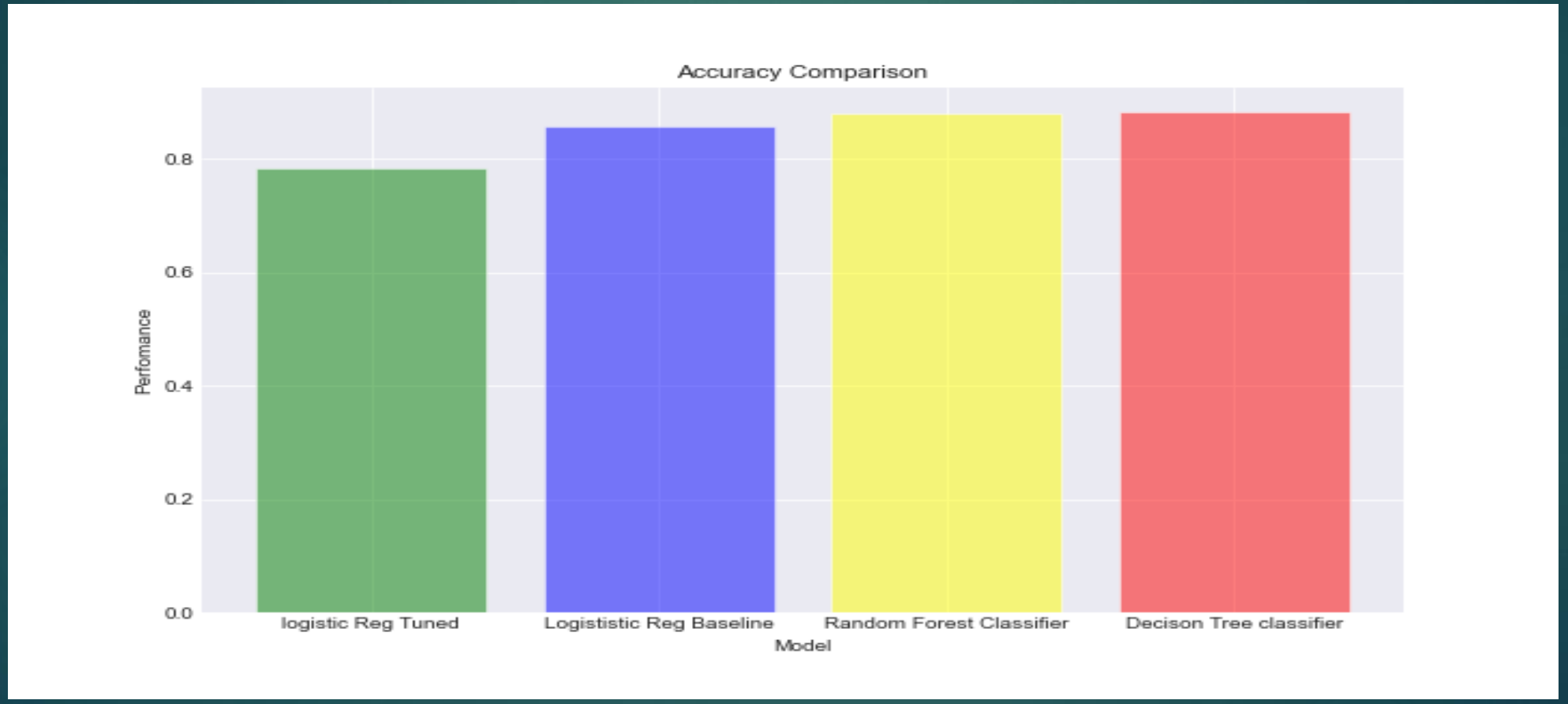
4.Random Forest Classifier

- Random forests Classifier models can handle high-dimensional datasets and are especially good at dealing with noise and outliers.
- The randomisation introduced during tree construction helps to reduce overfitting.
- Despite their complexity, random forests are interpretable due to the simplicity of individual decision trees.
- For this model, we will make use of the data that have their classes balanced by SMOTE.
- We can also get the metrics of the model as we did in the other models.

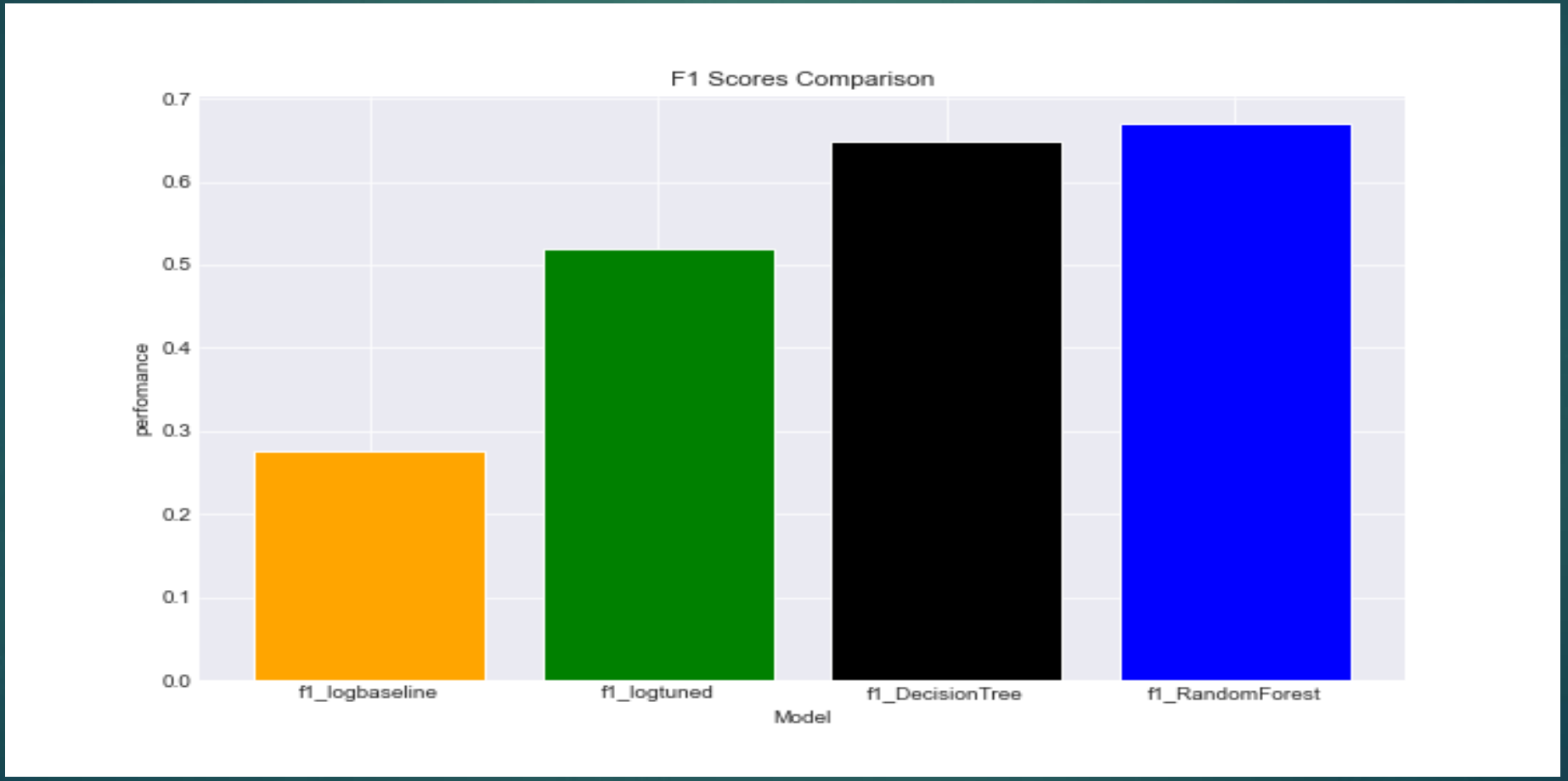
Evaluation of Models

- Now that we have four models, we need to settle on one that will be used by the SyriaTel communications company to predict customer churn with confidence.
- All the models have their metrics calculated and we can use them to compare the models.
- The metrics include; Accuracy, F1 score, recall, precission, AUC and ROC curves

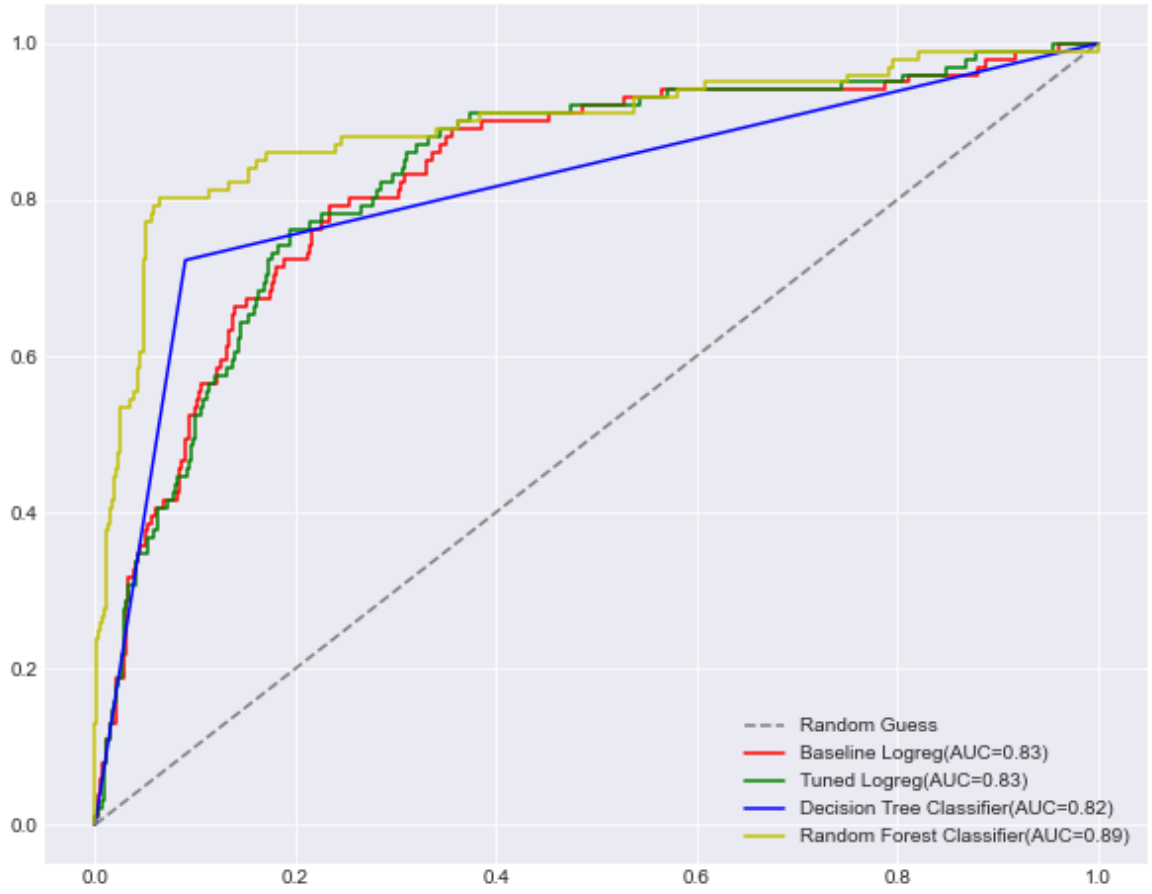
Accuracy Graph Comparison



F1 score Graph Comparison



ROC Curve Comparison



Conclusion

- the project shows that we can accurately predict customer churn on a significant level using a machine learning model.
- The comparisons indicate that the Random Forest Classifier Model is the most appropriate in our case being the one with good perfomance metrics overally.
- Random Forest Classifier is the best performing model with an ROC curve that is near the upper left corner of the graph, hence giving us the largest AUC (Area Under the curve).

Recommendations

- I would recommend that Syriatel make use of the Random Forest Classifier as the primary model for predicting customer churn. This model has a higher ROC curve and strong overall performance.
- I would recommend that the company focuses on tuning the call minutes and charges to best fit the customers. These efforts could include personalized offers or discounts on day charges.
- I would recommend, that Syriatel comes up with strategies to reduce on Customer Service calls. The more the customer service calls, the higher the likelihood of churn as this can easily irritate the customer.

Next Steps

- Explore more features such as Data Bundles plan that have not been captured by the dataset.
- Reach out to churned customers to understand their peronal reasons for leaving.
- Come up with ways to compete efficiently with other network providers.

Thank You!

ANY QUESTIONS?