Telecommunications Project

Done by: Kevin Kagia

Project Overview

SyriaTel is a telecommunications service provider that operates across the globe including USA among other countries. In the telecommunications industry, people often switch from one service provider to another. This change in taste and preference is what we refer to as churning. SyriaTel is no exception to this. Through their data, we can be able to understand the pattern behind customer churning and therefore SyriaTel and other companies in the industry will be able to anticipate customers who are likely to leave.

Project Overview

Business and data understanding

In order to carry out our task sufficiently, we first need an understanding of the business problem and to further understand the data that we will be using. The problem at hand is customer churning which has a negative outlook on the finances of the business. When customers keep leaving suddenly then you are unable to have an accurate financial long term plan for the business because without a forecast on the number of customers you'll retain or lose then you can't plan well. Therefore through developing a model, we can be able to anticipate future churning and therefore save the business a lot of money

Data Understanding

- **state**: The geographical location where the customer is residing.

- **account length**: The duration, measured in days, for which the customer has held their account.

- **area code**: The numeric code associated with the customer's phone number's geographical area.

- **phone number**: The unique phone number of the customer.

- **international plan**: A binary indicator for whether the customer has an international calling plan.

- **voice mail plan**: A binary indicator for whether the customer has a voice mail plan.

- **number vmail messages**: The count of voicemail messages the customer has sent or received.

- **total day minutes**: The total minutes the customer has spent on calls during the daytime.

- **total day calls**: The total number of calls the customer has made during the daytime.

- **total day charge**: The total monetary cost incurred by the customer for daytime calls.

- **total eve minutes**: The total minutes the customer has spent on calls during the evening.

- **total eve calls**: The total number of calls the customer has made during the evening.

- **total eve charge**: The total monetary cost incurred by the customer for evening calls.

- **total night minutes**: The total minutes the customer has spent on calls during the nighttime.

- **total night calls**: The total number of calls the customer has made during the nighttime.

- **total night charge**: The total monetary cost incurred by the customer for nighttime calls.

- **total intl minutes**: The total minutes the customer has spent on international calls.

- **total intl calls**: The total number of international calls the customer has initiated.

- **total intl charge**: The total monetary cost incurred by the customer for international calls.

- **customer service calls**: The count of calls the customer has placed to the customer service department.

- **churn**: A binary indicator for whether the customer has terminated their contract.

Contoso Suites

Modeling

SyriaTel is a telecommunications service provider that operates across the globe including USA among other countries. In the telecommunications industry, people often switch from one service provider to another. This change in taste and preference is what we refer to as churning. SyriaTel is no exception to this. Through their data, we can be able to understand the pattern behind customer churning and therefore SyriaTel and other companies in the industry will be able to anticipate customers who are likely to leave.
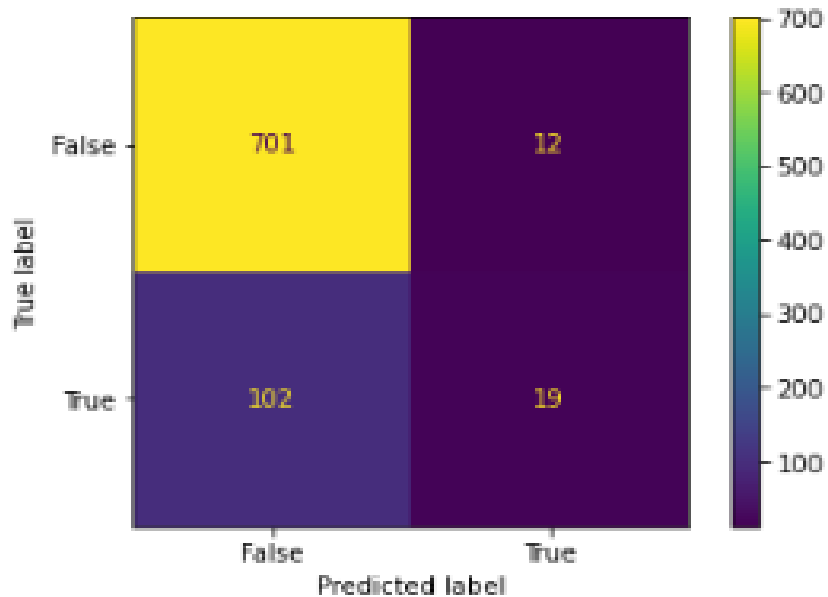
Project Overview

Logistic Regression

```
LogisticRegression(C=0.1, fit_intercept=False, solver='newton-cg')
```

```
Confusion Matrix:
 [[701  12]
 [102  19]]
```


Above, we have the tuned linear regression model. To the left, we have the performance of our model. The model has an accuracy of 86.3% meaning for every 100 data objects 86 of them will be correctly predicted therefore this is a very reliable model

```
0.0      0.863309
1.0      0.136691
dtype: float64
Precision: 0.6129032258064516
Recall: 0.15702479338842976
accuracy: 0.863309525179856
f1_score: 0.25
```

Logistic Regression

```
rmf = RandomForestClassifier(n_estimators=100, random_state=12)
rmf_model=rmf.fit(X_train, y_train)
```

```
0.0      0.902878
1.0      0.097122
dtype: float64
Classification Report:
              precision    recall  f1-score   support

         0.0       0.97      0.92      0.94       713
         1.0       0.63      0.81      0.71       121

    accuracy                           0.90       834
   macro avg       0.80      0.86      0.82       834
weighted avg       0.92      0.90      0.91       834


Confusion Matrix:
 [[655  58]
 [ 23  98]]
```
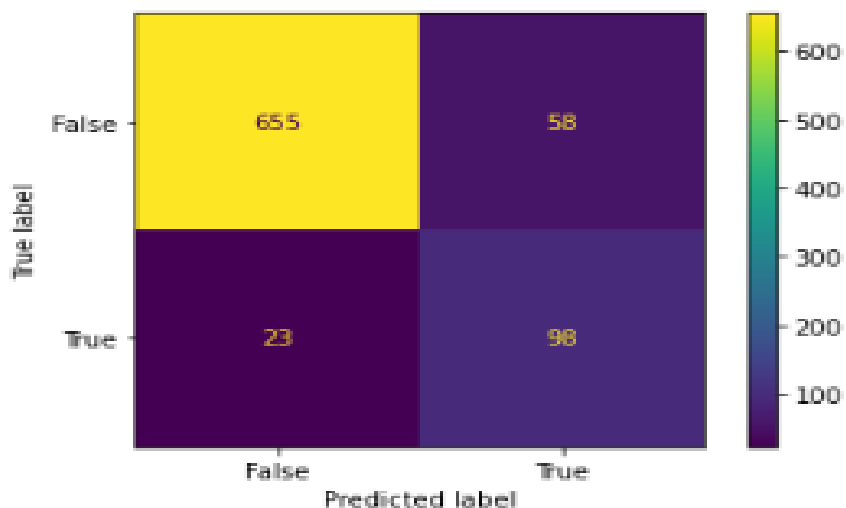
Above we have the code and the parameters we used to create our random forest model. To the left we have the performance of the model. As we can see the model has an accuracy of about 90% thus it is a very reliable model.

Further analysis

Here we have a ROC curve showing that the both models are quite efficient but the Random Forest is more effective though not so significantly

Receiver operating characteristic (ROC) Curve
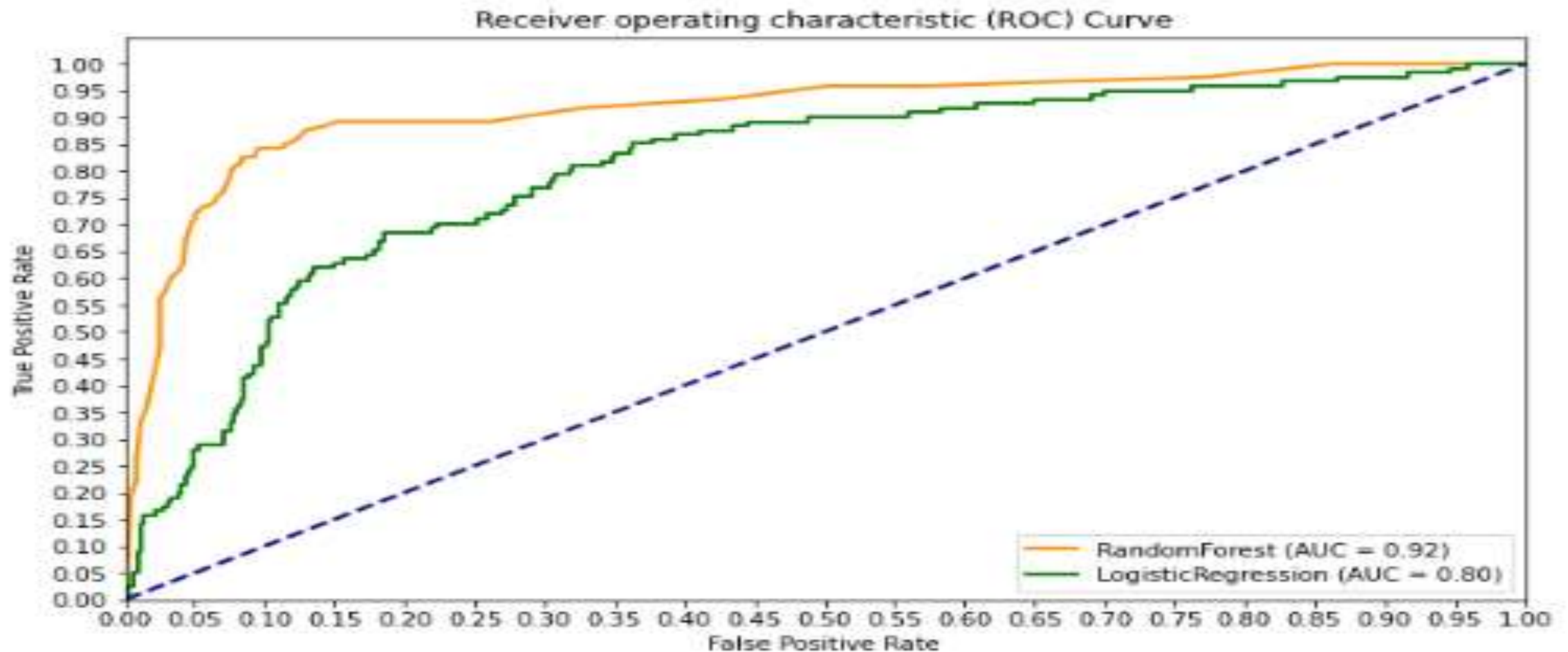
- **Model Suitability**: Both models are suitable for churning predictions, but Random Forest stands out with an AUC of 0.92 and a weighted average accuracy of 90%.

- **Customer Service Calls**: There is a clear correlation between the number of customer service calls and the likelihood of churn. The more customer service calls a customer makes, the more likely they are to churn.

- **Calling Charges Impact**: Reducing calling charges is expected to have a positive effect on reducing the churn rate. Lower charges may make customers less likely to terminate their contracts.

- **International Users**: Interestingly, international users appear to have a lower incidence of churning. This suggests that international plan subscribers may be more loyal or satisfied with their service.

Conclusions

- **Model Selection**: Given that Random Forest outperformed other models with an AUC of 0.92 and a 90% weighted average accuracy, it is recommended to prioritize the use of Random Forest for churning predictions.

- **Customer Service Enhancement**: To reduce churn, consider enhancing the customer service experience, as a higher number of customer service calls appears to correlate with increased churn. This could involve improving support, response times, or providing additional support resources.

- **Pricing Strategy**: To mitigate churn, consider implementing a pricing strategy that includes reducing calling charges. Lower charges can make the service more cost-effective for customers and potentially reduce the likelihood of contract terminations.

- **Promote International Plans**: Given that international users have a lower incidence of churning, consider promoting international calling plans. This could include marketing and incentivizing international plans to retain more customers and increase loyalty among subscribers.

Recommendations

Thank You