



CUSTOMER CHURN PREDICTION

By CALVINCE KAUNDA

Business Problem and objectives

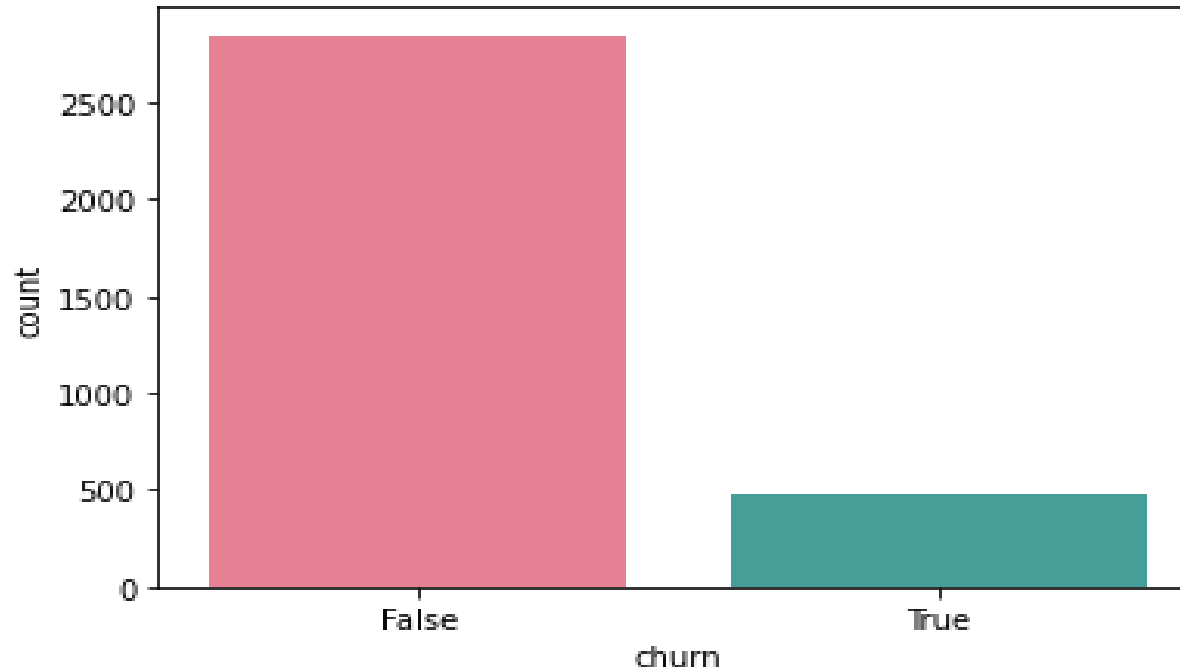
- ▶ SyriaTel is a telecommunications company that is interested in a model that predicts whether a customer churns or not
- ▶ Our task is to analyze the dataset and come up with a model that best predicts churn based on a variety of factors
- ▶ The model is based on this Dataset : [Churn in Telecom's dataset](#) from Kaggle.

Data understanding and exploration

- ▶ Analysis was done using the pandas library and uncovered some important features.
- ▶ The dataset had 3333 rows and 21 columns
- ▶ Not all columns are necessary for our analysis and unnecessary columns were dropped
- ▶ Visualizations were also plotted using the Seaborn and Matplotlib Libraries to better understand relationships and distributions.
- ▶ The dataset did not have any missing values

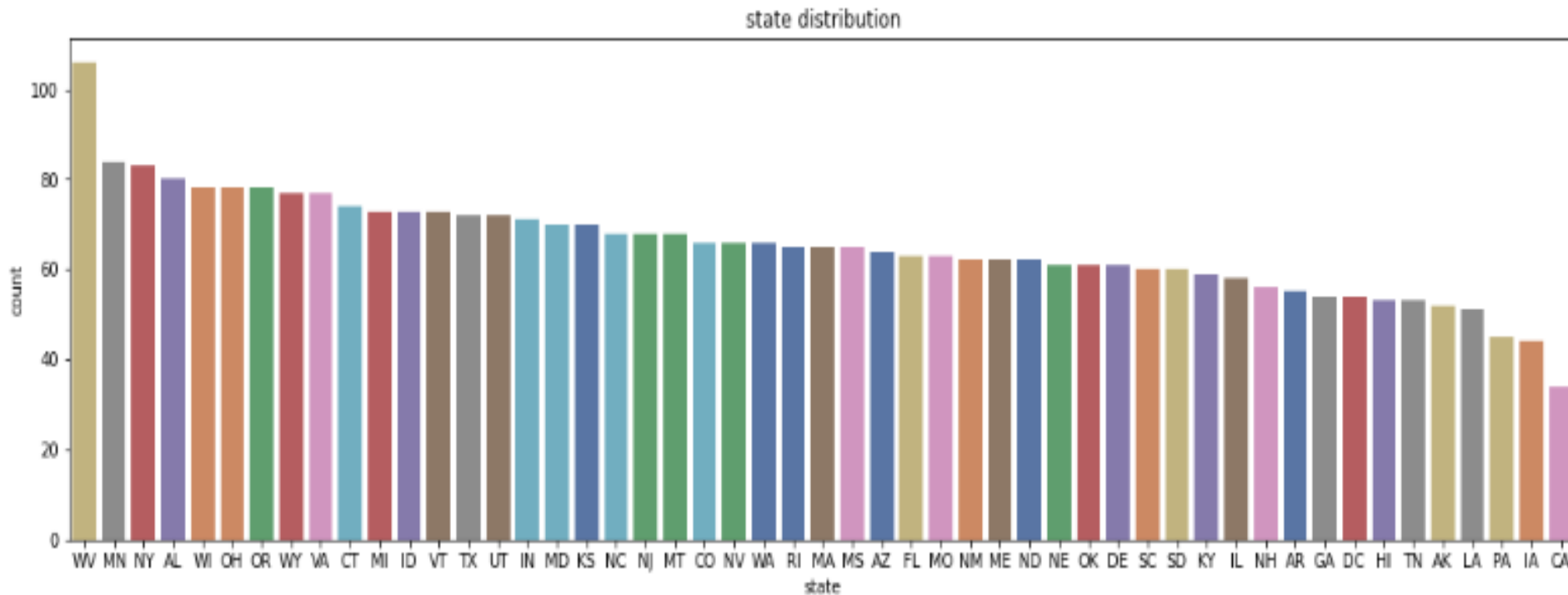
Exploratory Data Analysis

- **Churn Distribution:** We had 85% customers who did not churn and 14.5% of customers who churned.



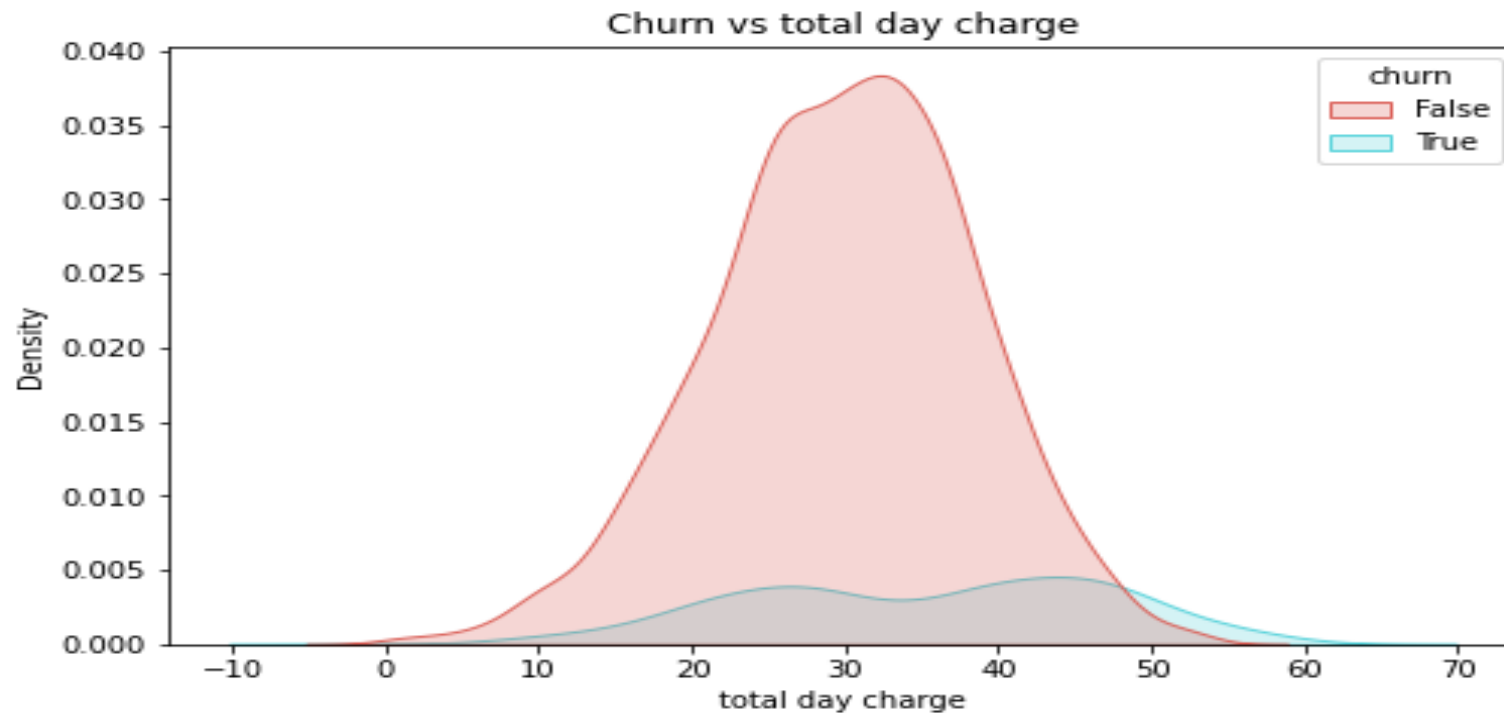
Exploratory Data Analysis

- **State Distribution:** Majority of our customers are from West Virginia, Minnesota, New York, Alabama and Wisconsin



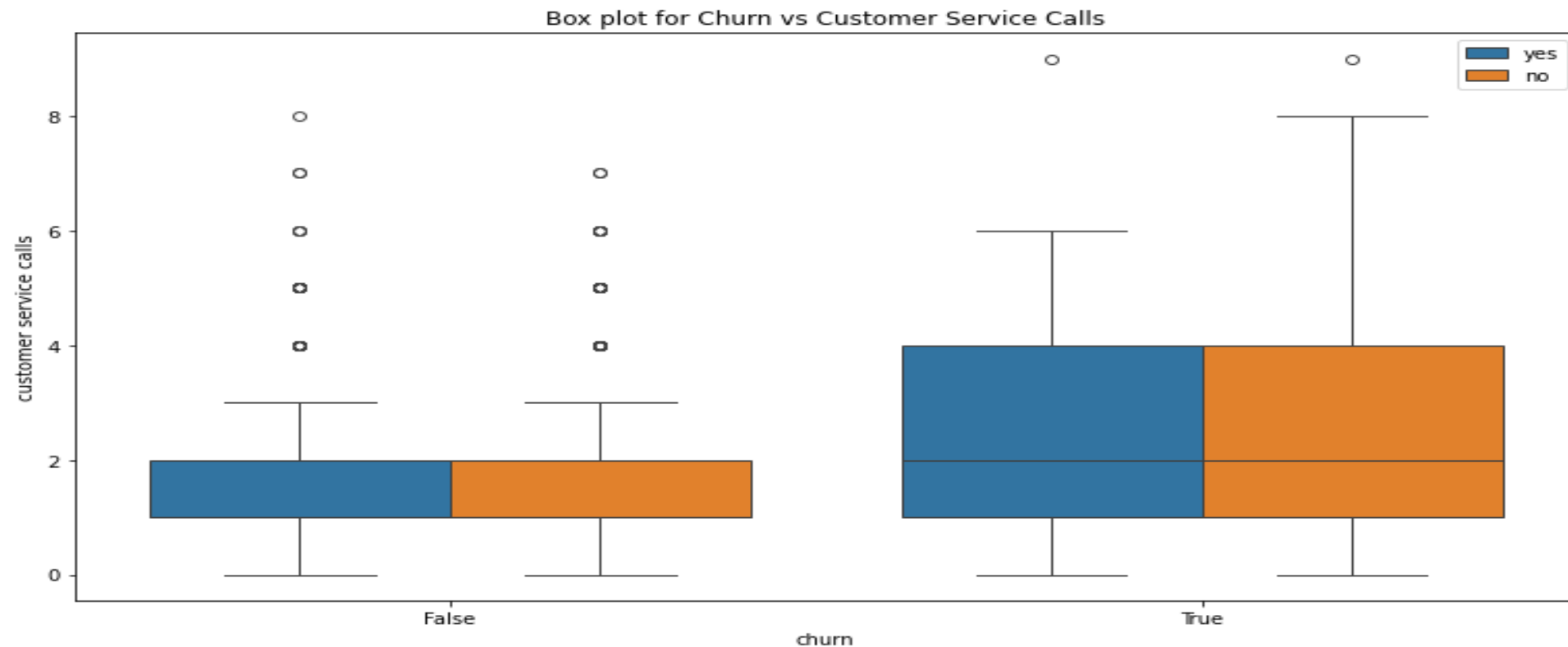
Exploratory Data Analysis

- Customers with **HIGHER TOTAL DAY CHARGES ARE MORE LIKELY TO CHURN**



Exploratory Data Analysis

- More customer service calls, higher risk of churn



Data preprocessing

- ▶ Scikit Learn was a key python library used to preprocess the data and generate a model.
- ▶ The library comes with multiple tools. A few of them include train-test-split, ordinal encoding and the MinMaxScaler
- ▶ Techniques such as Ordinal Encoding were used to encode data and convert categorical features
- ▶ The Dataset was split using the train-test-split where the dataset was split into a train size of 85% and a test size of 15%
- ▶ Scaling was also performed using MinMaxScaler to convert the whole dataset to fall in between 0 and 1

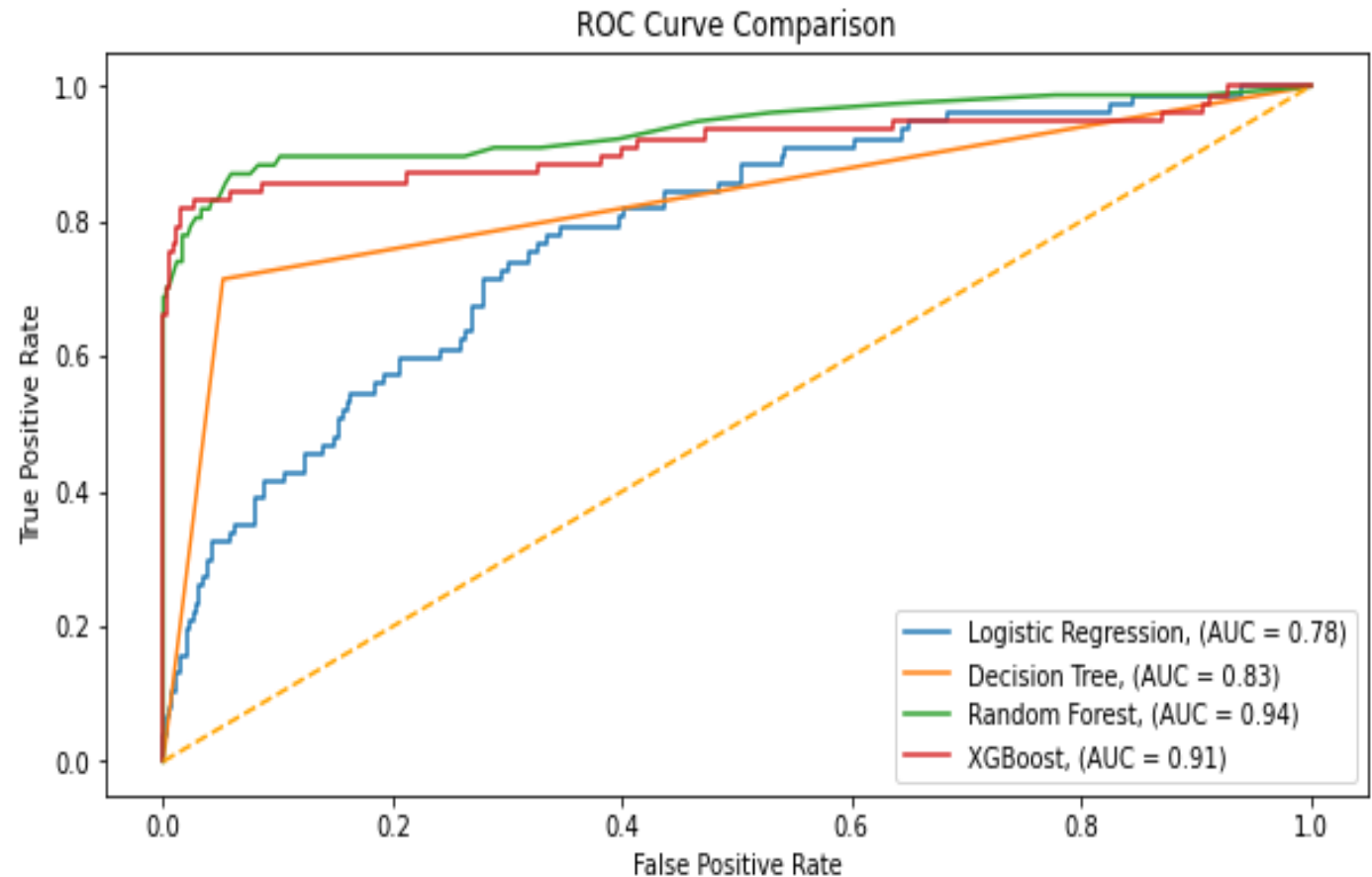
Modelling

- ▶ 4 models were trained on the data and tested on the data
- ▶ They include:
 1. Logistic Regression Model
 2. Decision Tree Model
 3. Random Forest Model
 4. XGBoost Model.
- ▶ The dataset was imbalanced and class weights used to balance the data

Model Evaluation: ROC-AUC Curves

► The best models before tuning were as follows:

1. Random Forest
2. XGBoost
3. Decision tree
4. Logistic Regression

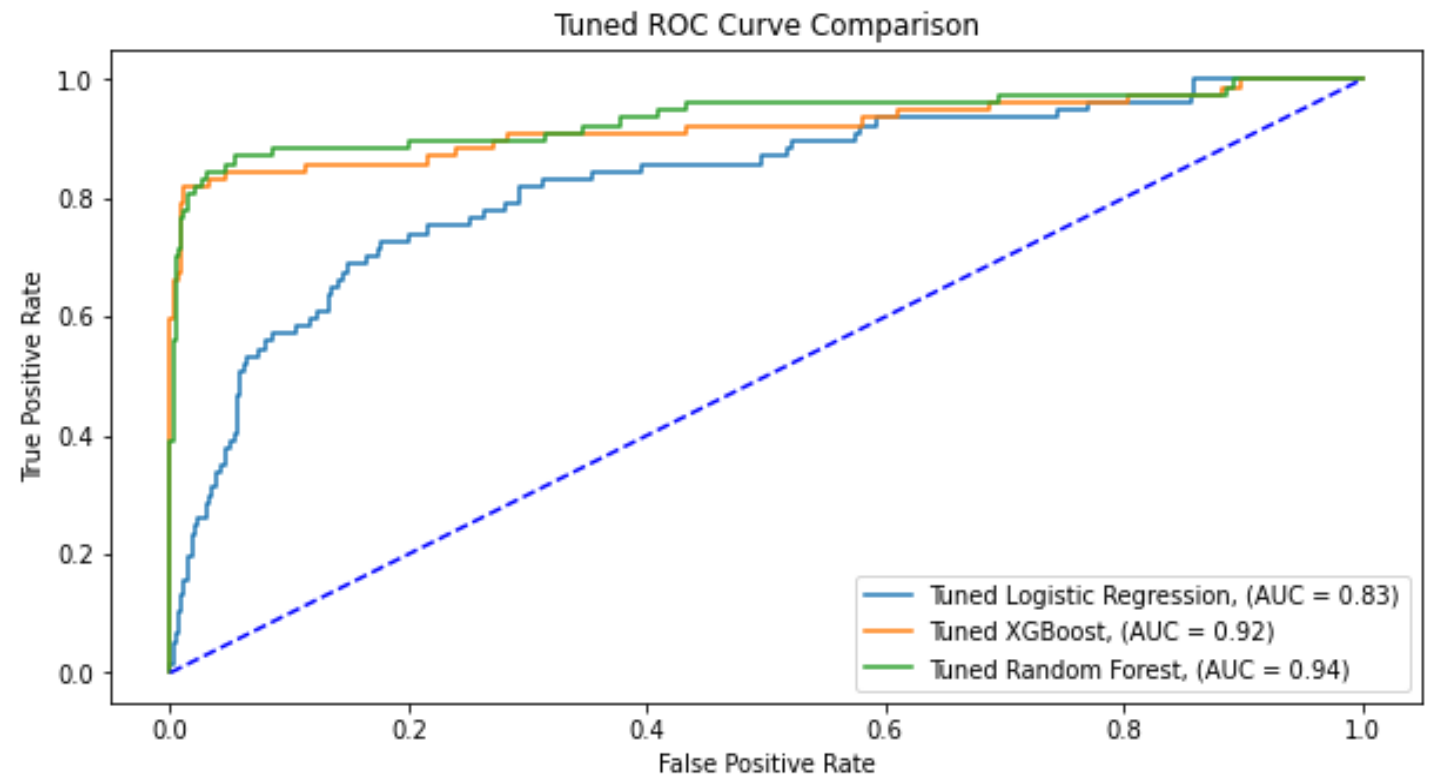


Parameter Tuning

- ▶ Random Forest, XGBoost and Logistic Regression models were tuned to improve their overall scores
- ▶ After tuning the Random Forest model did not show any improvement in score but instead showed signs of SLIGHT OVERFITTING.
- ▶ The Logistic Regression Model did improve in performance but its performance was not as good as compared to the Random Forest and XGBoost models

Model Evaluation after tuning: ROC-AUC Curves

- The following were the rankings after parameter tuning:
1. XGBoost Model
 2. Random Forest Model
 3. Logistic Regression Model



Findings

- ▶ The best model was the tuned XGBoost which showed good generalization on even unseen data
- ▶ XGBoost model could still be improved further by using feature engineering
- ▶ Random Forest Model also showed great potential but showed signs of SLIGHT OVERFITTING. Further tuning and feature selection can reduce this
- ▶ **Recommended model to be deployed – XGBoost Model**

Conclusion and insights

- ▶ Customer service quality should be improved in order to reduce the number of customer service calls. Training programs to the customer care team can be applied to have the team provide effective solutions to issues affecting customers.
- ▶ Discounts should be introduced as customers with higher charges are more likely to churn. Discount tactics such as discounting customers who have consistently high charges could help reduce churn.
- ▶ Voicemail plans should be marketed more to try and increase the adoption of voicemail plans as frequent usage of voice mail plans shows that the customer is less likely to churn. New offers such as discounting all customers who have voice mail plans could lead to more customers adopting the voice mail plans

THANK YOU

Any Questions?

CONTACTS:

Calvince Kaunda

Visit my [Github](#).