



DATA
SCIENCE

SYRIATEL CUSTOMER CHURN PREDICTION PROJECT

Presented by: Sandra Jebet Kiptum

Introduction

customer churn refers to the phenomenon where customers stop using a company's services. For Syriatel, a leading telecom provider, retaining customers is crucial for sustaining revenue and growth. This project aims to predict customer churn, enabling Syriatel to identify at-risk customers and implement strategies to retain them.

Customer churn prediction is vital for several reasons:

- Cost efficiency - it is cheaper to retain customers than to acquire new ones
- Revenue stability
- Customer satisfaction
- Competitive advantage



Objective



Building a model

The core objective of this project was to build a machine learning model that will predict if a customer will churn or not



Why?

the importance of the model is that it will allow the company to know a customer is leaving even before they do, giving them a chance to take preventive measures



Data preprocessing



01

the data was acquired from syriatel, then i later acquired it from kaggle



02

For data processing we checked for duplicates, null values , outliers and so much more. the data was pretty clean apart from a few things here and there which was quickly handled



03

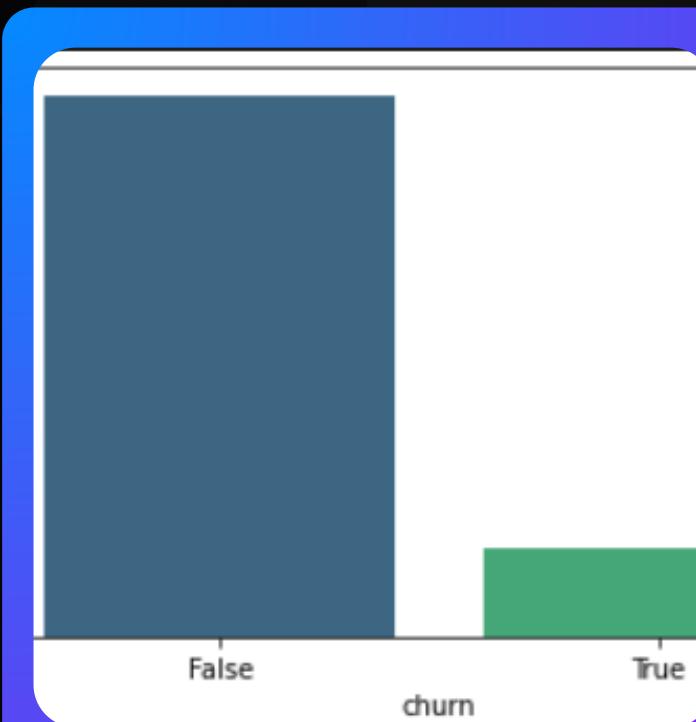
next was EDA where we saw that customer service calls were easily the best indicator for churning. the more the calls the more the likelihood of churning



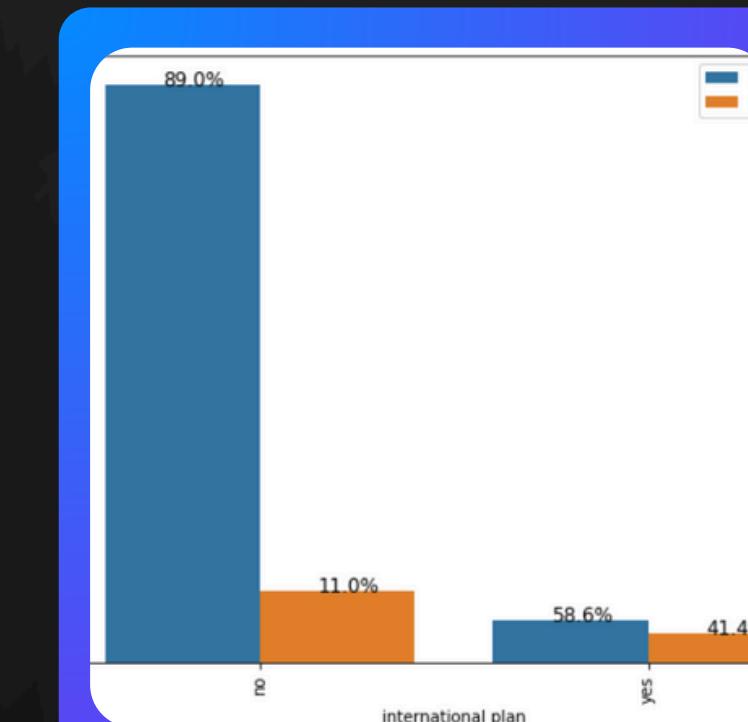
04

lastly was the preprocessing before modelling. we scaled numerical data, one hot encoded categorical data, merged the two, split into train and test the used smote to resample cause of the class imbalance

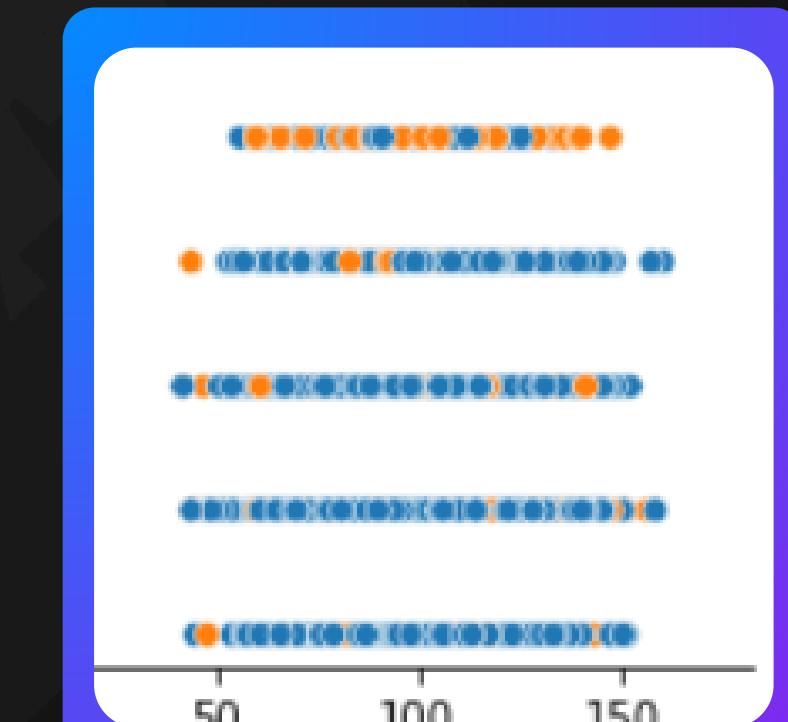
Exploratory Data Analysis



we saw that though the number of churners was less than that of non-churners it was still a scary number



we also noticed that customers with international plans were most likely to churn and the company should look into the short comings of the plan



we also noticed that the more the customer service call the more the likelihood of churning



Modelling approach

the approach used for modelling was basically trial and error.

i moved from one model to the other comparing their metric trying to find the best one.

the metrics (F1 score, precision, accuracy and recall) we compared between test and train datasets as well as a cross validation

Models used



Logistic Regression

cross validation

Precision: 0.73 | Recall: 0.70 | Accuracy: 0.77

train data

Precision: 0.75 | Recall: 0.72 | Accuracy: 0.79 F1-Score: 0.73

test data

Precision: 0.41 | Recall: 0.66 | Accuracy: 0.82 F1-Score: 0.51



K Nearest Neighbours

cross validation

Precision: 0.86 | Recall: 0.98 | Accuracy: 0.93

train data

Precision: 1.0 | Recall: 1.0 | Accuracy: 1.0 F1-Score: 1.0

test data

Precision: 0.43 | Recall: 0.48 | Accuracy: 0.83 F1-Score: 0.46



Decision Tree

cross validation

Precision: 0.92 | Recall: 0.83 | Accuracy: 0.90

train data

Precision: 0.95 | Recall: 0.84 | Accuracy: 0.92 F1-Score: 0.89

test data

Precision: 0.75 | Recall: 0.8 | Accuracy: 0.93 F1-Score: 0.77

Models used

Random Forest

cross validation

Precision: 0.94 | Recall: 0.89 | Accuracy: 0.93

train data

Precision: 0.99 | Recall: 0.94 | Accuracy: 0.97

test data

Precision: 0.77 | Recall: 0.7 | Accuracy: 0.93



XG-BOOST

cross validation

Precision: 0.97 | Recall: 0.92 | Accuracy: 0.96

train data

Precision: 1.0 | Recall: 1.0 | Accuracy: 1.0 F1-Score: 1.0

test data

Precision: 0.9 | Recall: 0.8 | Accuracy: 0.96 F1-Score: 0.85



Modelling

This was the final metric that i used to compare the models; AUC values and ROC curve

92%

Random Forest

91%

XG-BOOST

90%

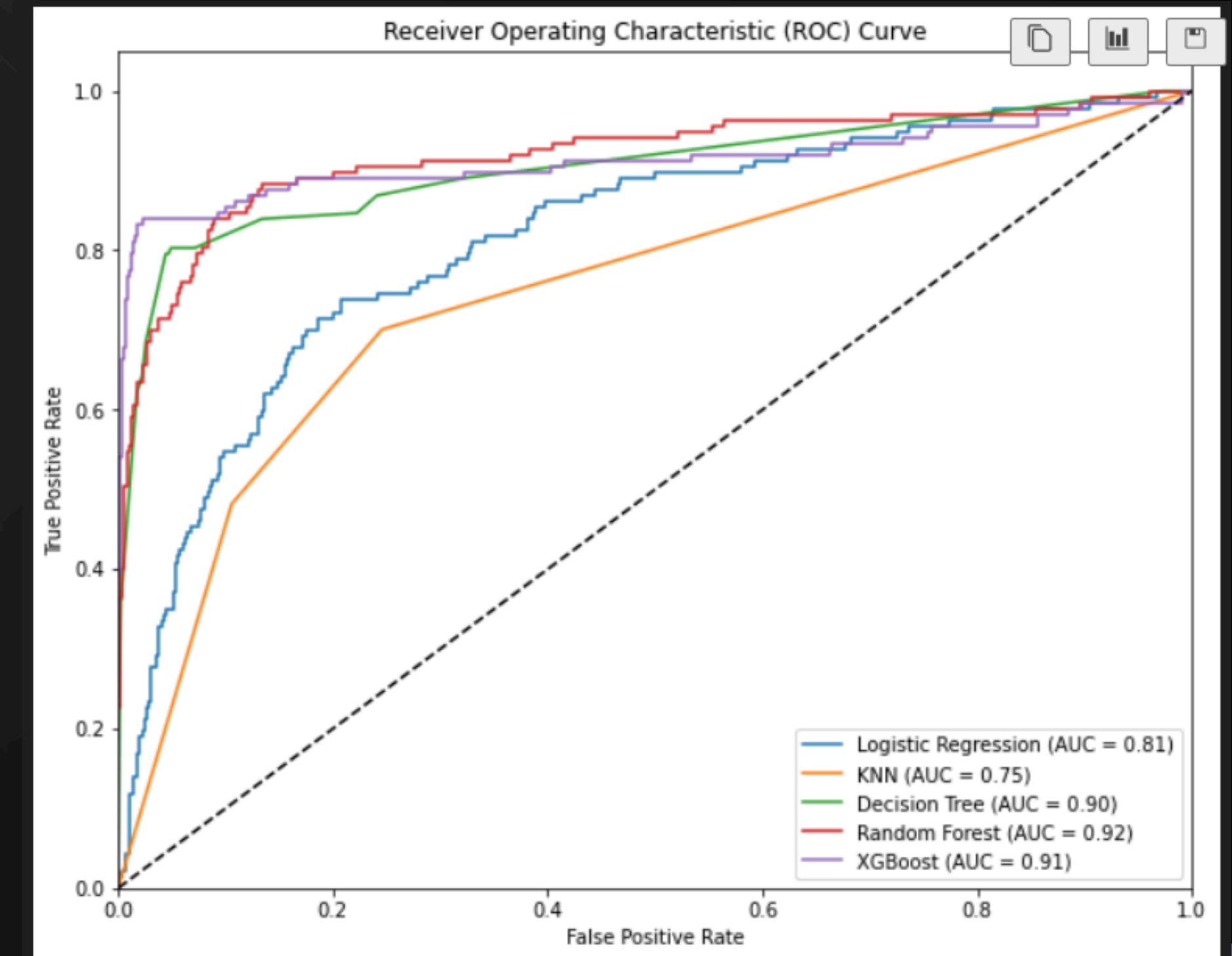
Decision Tree

81%

Logistic Regression

75%

KNN





Recommendations

- i would reccomend that SyriaTel look into using the XG-BOOST model to predict if a customer will soon churn and find ways to prevent it even before it happens
- i would also reccomend that syriatel gets developers to get developers to build a platform to allow employess access this model easily



Data
Science

THANK YOU