



SYRIATEL CUSTOMER CHURN PREDICTION

BY CELIAJOY OMIAH

OVERVIEW



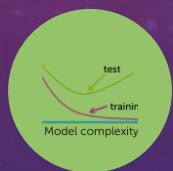
Business Understanding



Data Understanding and Analysis



Modeling



Evaluation and Recommendation



OVERVIEW

- Syriatel is a telecommunication company founded in 2000 based in Syria with its headquarters in Damascus. It offers GSM, 3G, and 4G network services.
- Syriatel is faced with a challenge where there is a high churn rate with most of its customers abandoning their services and switching to their competitors.

Problem statement

- To develop a machine learning model that can be used to predict and determine the probability of a customer churning a service and what factors result in this.

OBJECTIVES

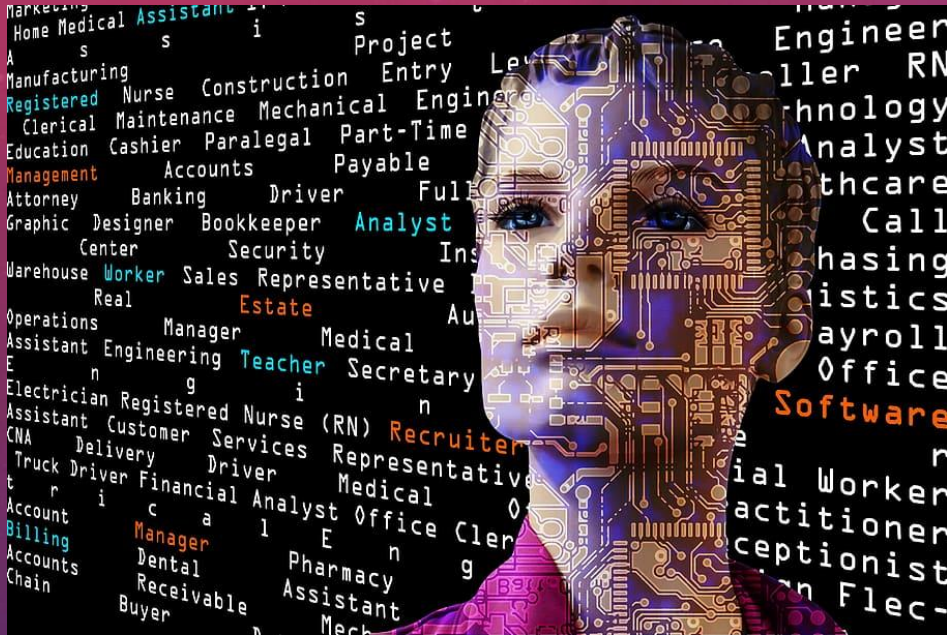


- To identify the features that are likely to cause a customer to churn.
- To develop a machine learning model that will accurately predict the customers that are at a high risk of churning.
- To come up with measures to prevent customers who are at high risk from churning

BUSINESS UNDERSTANDING

- Customer attrition, also known as churn, refers to the loss of clients or customers and is particularly pertinent to industries such as telephone service companies, internet service providers, TV companies, and insurance firms. Managing and mitigating customer attrition is a crucial concern for businesses in these sectors.

DATA UNDERSTANDING



- The data used has been sourced from Kaggle. The dataset contains a record of 3333 rows and 21 columns containing different features. Out of the 21 columns, 4 are categorical and 17 are numerical.

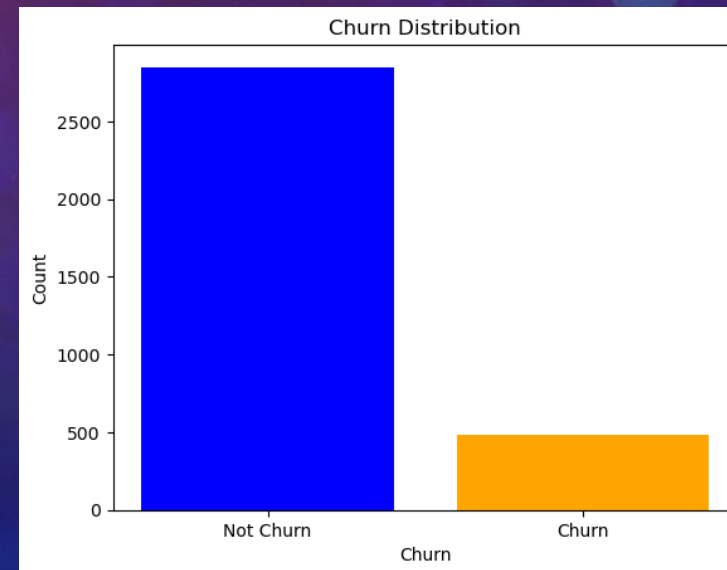
DATA ANALYSIS

Data Cleaning

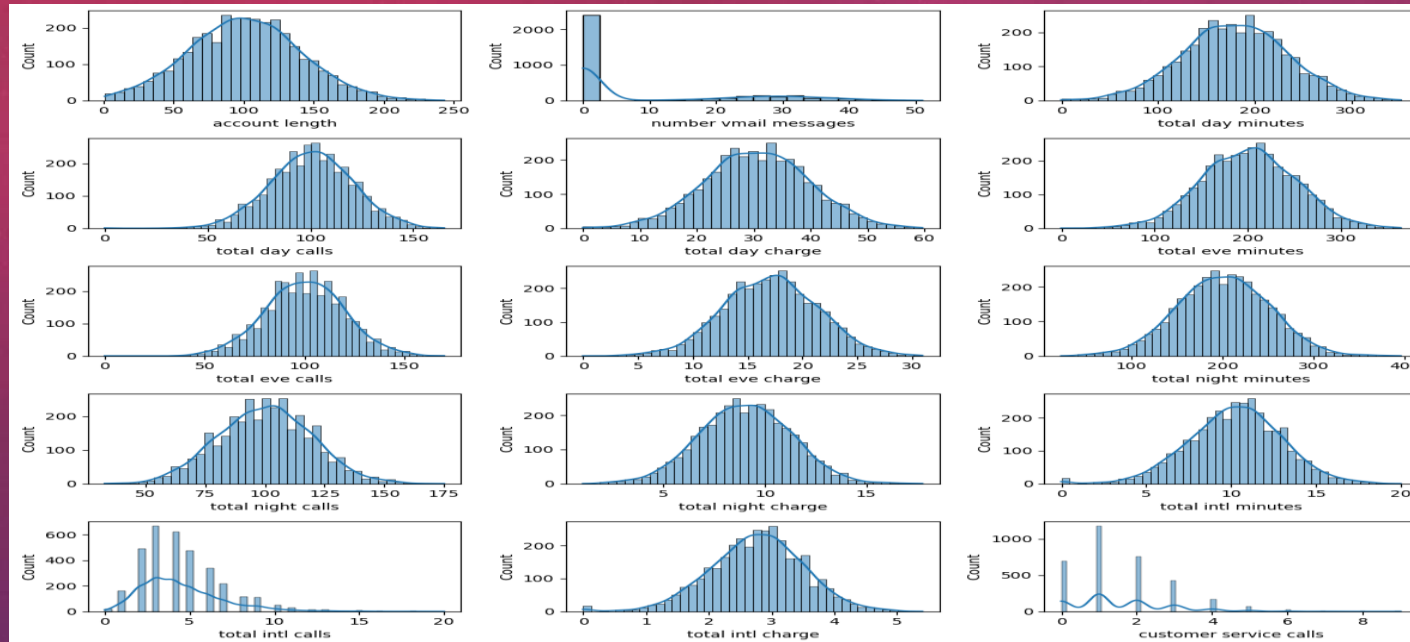
- We went through the dataset and inspected if there were missing values and null values. From this process there were no missing values and null values. This shows the dataset we are working with to be complete. There were no duplicates in the dataset. The column representing the phone number was dropped because it was not important for our modelling.

Univariate Analysis

- We started our analysis by looking into each of the features in our dataset starting with the target variable and visualizing it to see the distribution.
- From the graph above about 483 customers out of the total 3333 in the dataset have churned (meaning terminated) their contracts with the telecommunication company.



A histogram showing the distribution of all the numerical features enabled us to see most of the features having a normal distribution curve. The analysis is shown below



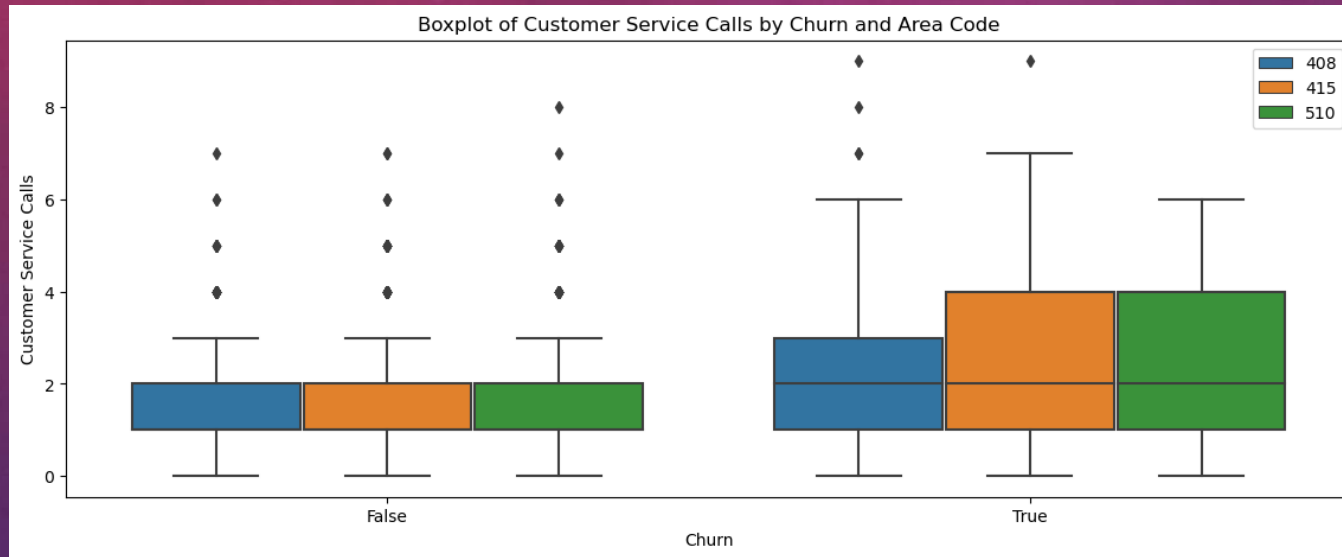
1. The distribution in the Account length is positively skewed, indicating that the majority of customers have shorter account lengths. This suggests potential inaccuracies in predicting customer churn, as longer account lengths might be a more accurate predictor.

2. The distribution of the Number of Voice Mail Messages is relatively even, indicating that the number of voicemail messages is not a particularly strong predictor of customer churn.

3. Total Day Minutes, Total Day Calls, and Total Day Charge: These features show a significant amount of potential noise in the data, as the distribution is positively skewed and there is a significant amount of overlap between the classes. This is just a brief explanation of the distribution of the features,

Bivariate Analysis

- We analyzed two features to explore their relationship and see how the changes in each feature affect the other. In the boxplot below we can conclude that most of the customers who have terminated their contract are from the area code 415 and 510.



MODELING

The following algorithms were employed:

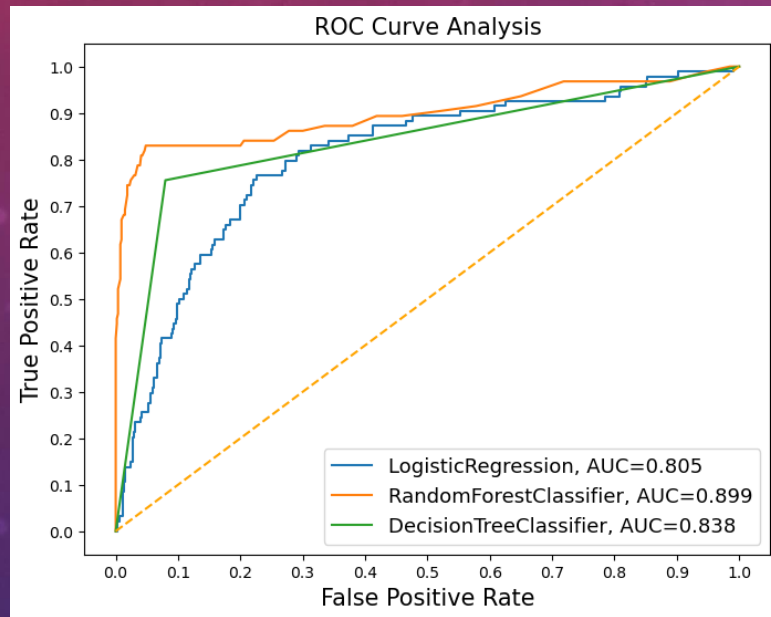
1. Logistic Regression
2. Decision Tree
3. Random Forest

Metrics used

- Recall – which focuses on the model's ability to correctly identify positive cases
- ROC-AUC Curve –plots the true positive rate against the false positive rate at various classification thresholds.

EVALUATION

ROC Curve Analysis



Recall results

- Logistic regression -0.76
- Decision tree - 0.76
- Random forest -0.79

The tuned Random Forest model achieved a recall score of 0.79 and an ROC_AUC of 0.899 which is an improvement.

CONCLUSION



- In summary, while the Random Forest classifier has achieved satisfactory results, the commitment to continuous improvement prompts a deeper dive into feature engineering. This pursuit aims to refine the model's predictive capabilities and potentially elevate the recall score, contributing to an even more effective and robust customer churn prediction system.

RECOMMENDATION

- In response to the observed higher churn rates in area codes 415 and 510, it is recommended to implement targeted promotional strategies, specifically offering discounts and promotional offers to customers within these areas. By providing incentives, such as exclusive discounts, the aim is to foster customer loyalty and discourage churn. This localized approach acknowledges the unique characteristics and challenges associated with these specific geographical regions.
- Additionally, recognizing the critical role of customer service in customer retention, there is a proposal to invest in improving the quality of customer service. This involves enhancing the training programs for customer service representatives, equipping them with the skills and knowledge to promptly and effectively address customer issues. The objective is to elevate customer satisfaction levels and, consequently, reduce the likelihood of churn. By minimizing the need for customers to reach out with concerns, this initiative contributes to a smoother overall customer experience.
- An evaluation of the existing pricing structure for day, evening, night, and international charges is recommended. Considering the impact of charges on customer decisions to churn, there is a suggestion to explore adjustments to pricing plans or the introduction of discounted packages. This strategic pricing review aims to address any discrepancies and make the

thank
you