

b-irungu / PHASE-3-BINARY-CLASSIFICATION-PROJECT Public

0 stars 0 forks

Star

Watch

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

main

b-irungu Add files via upload ...

18 minutes ago 22

[View code](#)

README.md



SYRIATEL COMPANY CUSTOMER CHURN ANALYSIS PROJECT

USING BINARY CLASSIFICATION TO BUILD A MODEL THAT ACCURATELY PREDICTS CUSTOMER CHURN TO HELP SYRIATEL COMPANY IDENTIFY THE FACTORS CONTRIBUTING TO THE CHURN AND TAKE PROACTIVE ACTIONS TOWARDS RETAINING THEIR CUSTOMERS

REPOSITORY OUTLINE

This repository contains the following

- A jupyter notebook
- A non technical presentation.

- A CRISP-DM report
- A README file
- SyriaTel Customer churn data
- SyriaTel company logo
- churn distribution visual

OVERVIEW

This is a binary classification project that conducts a churn analysis for SyriaTel which is a telecommunication company in Syria. The project aims at identifying the factors that contribute to customer churn and Develop a classifier that predicts which customers are likely to churn to enable SyriaTel take appropriate actions and reduce customer attrition.

DATA

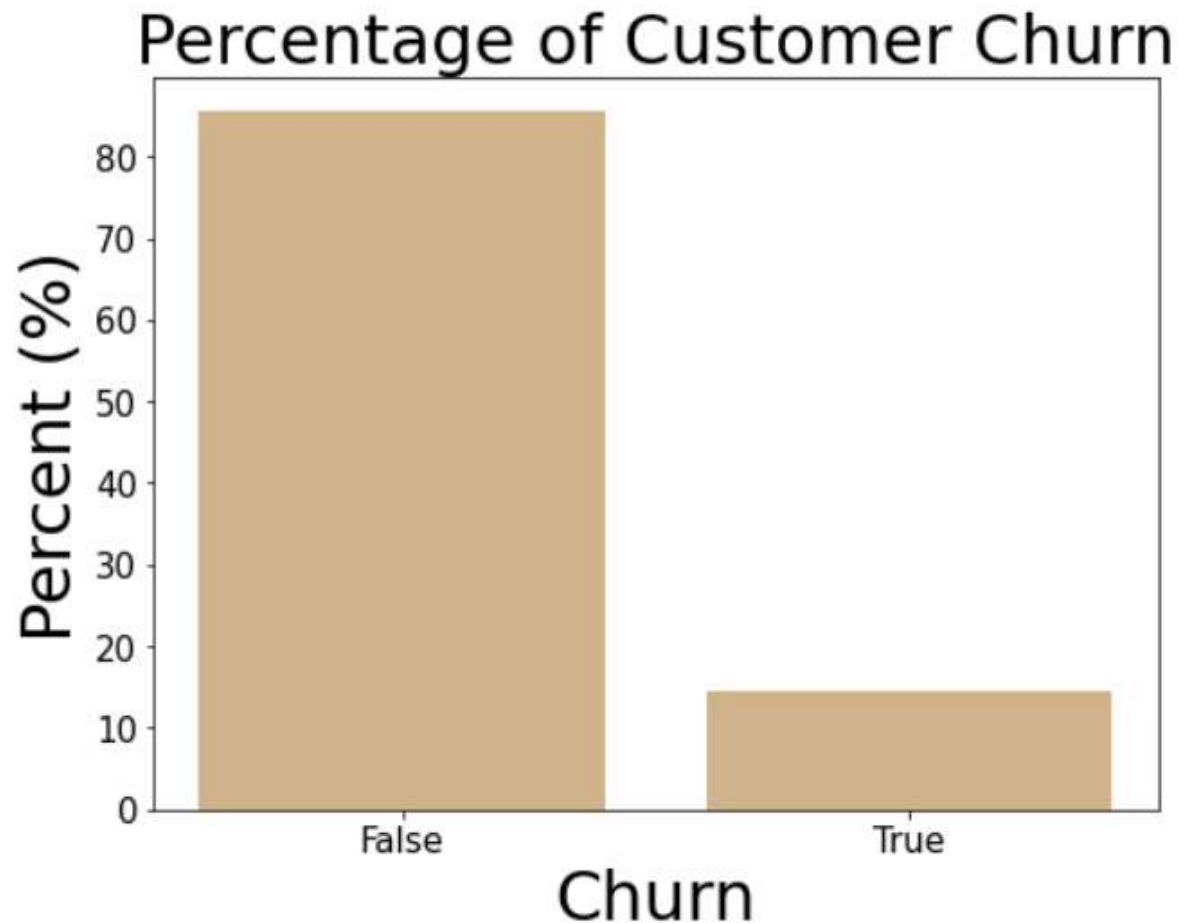
This project uses the SyriaTel Customer churn data set which is retrieved from Kaggle: <https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset>. The dataset is suitable for this project based on the following aspects;

- It contains comprehensive information on various features of the telecommunication services provided by the company including account length, area code, phone number, international plan, voice mail plan, and various other features related to customer usage patterns.
- The data set also contains the feature 'churn,' which indicates whether a customer has churned or not which will be used as the target variable making the dataset suitable for analyzing customer churn as it contains all the necessary information to develop a predictive model.
- The data he dataset is large enough, with over 3,000 records, to support the development of a reliable and accurate predictive model.

Target and predictor variables

The data has various features where `churn` was the target variable feature while the rest were the predictor variable features. the predictor variable features include the `state` , `account length` , `area code` , `phone number` , `international plan` , `voice mail plan` , `number vmail messages` , `total day minutes` , `total day calls` , `total day charge` , `total eve minutes` , `total eve calls` , `total eve charge` , `total night minutes` , `total night calls` , `total night charge` , `total intl minutes` , `total intl calls` , `total intl charge` and `customer service calls`

Customer churn distribution



METHOD

The project used the CRISP-DM data science process to analyse and create a model best for the churn analysis. All the data science steps are described in the CRISP-DM Report:<https://github.com/b-irungu/PHASE-3-BINARY-CLASSIFICATION-PROJECT/blob/main/FINAL%20CRISP-DM%20DATA%20REPORT%20-.pdf>. The projects used the following Binary classifiers

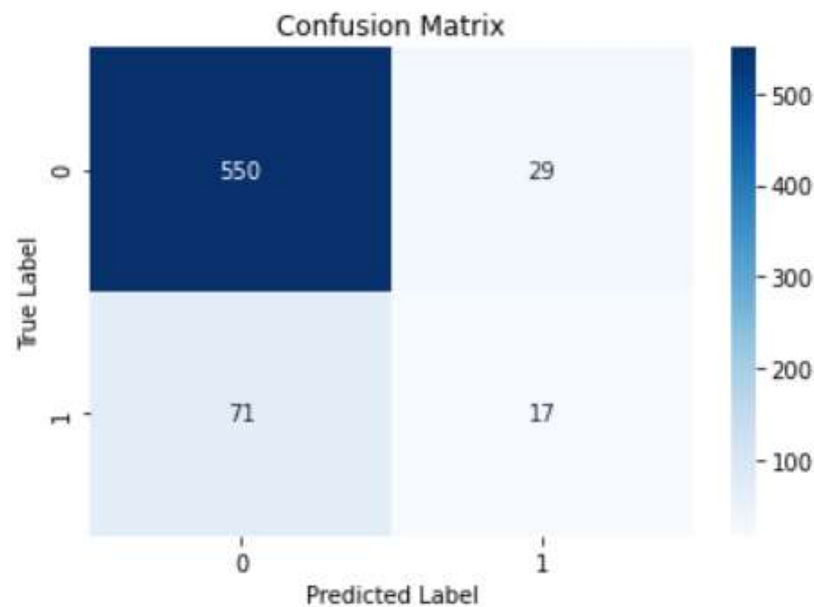
- KNN(K-Nearest Neighbours)
- Logistic Regression
- Decision Trees
- Random Forest

FINDINGS

The findings of the various classifiers used in this project are as follows:

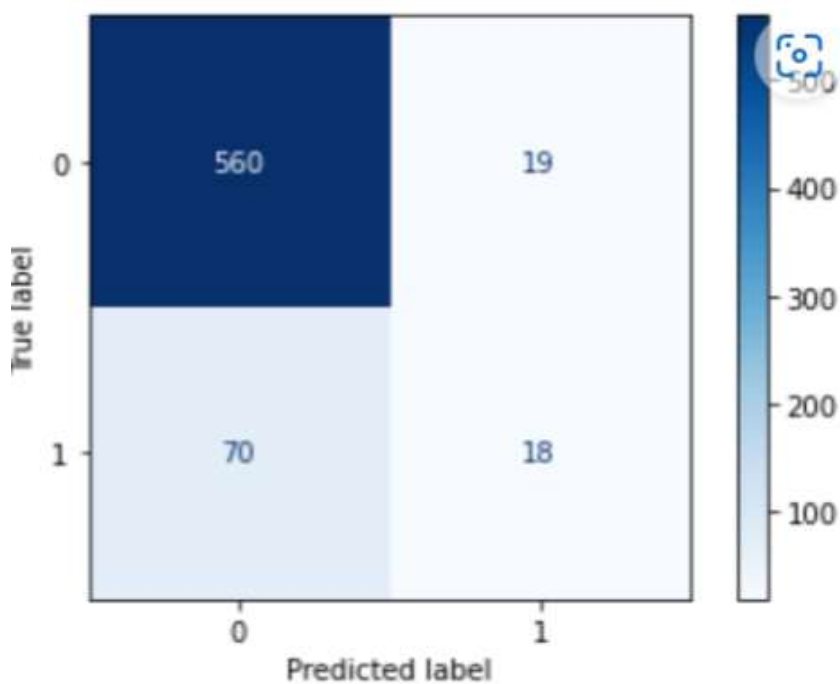
KNN findings

This classifiers had the lowest performance(based on the recall evaluation), as the recall score ws 19.31%. with this score the model had a high rate of producing false negative churns which would result to mislead decision making if the same was adopted.



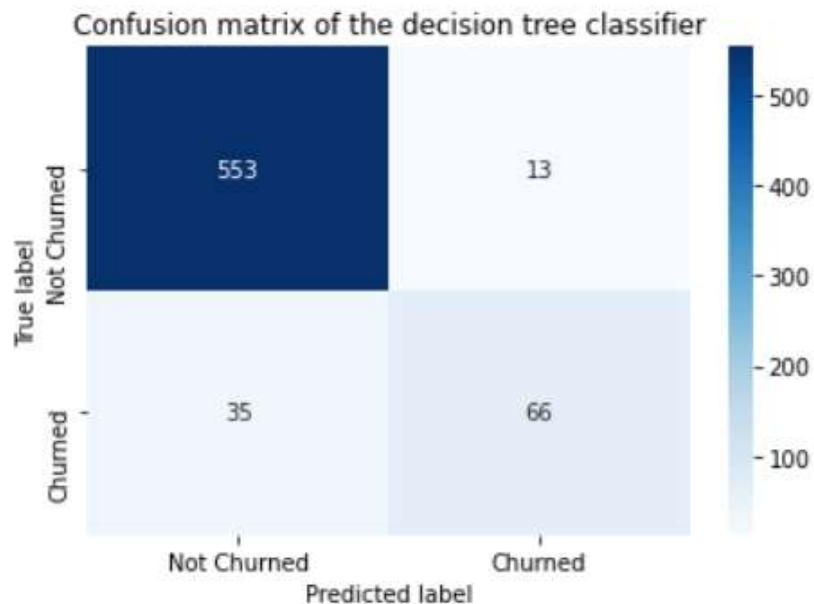
Logistic Regression findings

The recall score for the classifier was 28% which also indicated that the classifier performed poorly on correctly predicting customer churn, thus the classifier could not be adopted



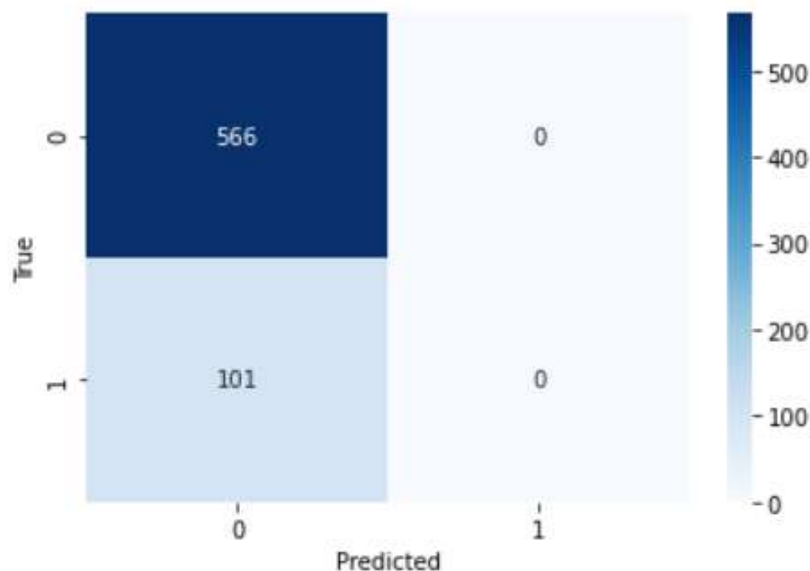
Decision Trees findings

The model reported a high accuracy of 0.929 and the highest recall score of 78%. with the highest recall and accuracy, the model was adopted and business recommendations and conclusions drawn from it.



Random Forest Findings

Random Forest had a Recall Metric of 23%. 23% of the predictions are False Negatives which means that the model will 23% of the time predict that a customer will not churn yet the customer churns.



CONCLUSION

Based on the findings, the business conclusion can be drawn as follows:

- Importance of Recall: In the context of predicting customer churn, the focus was placed on optimizing for Recall. By prioritizing Recall, the goal was to minimize the number of customers who are incorrectly classified as non-churners.
- Best Model: Among the models explored, the decision tree Classifier performed the best since was able to correctly identify 78% of the customers who were likely to churn.the model had an accuracy score of 92.8%

- The factors that mostly influence churn of customer include total day charge, customr service calls and number oof voice mail messages.
- Predicting customer churn is an ongoing process, and it is important to continuously refine and improve the model. Regularly monitoring the model's performance, collecting new data, and incorporating feedback from business stakeholders can lead to better predictions and more accurate identification of customers who are at risk of churning

BUSINESS RECOMMENDATION

Determine the unique needs of the following customers and meet them;

- The company should ensure continoous prediction of the factors influencing churn by contionously collecting new data and improving the model as prediction is an ongoing process.
- To reduce customer churn, the company should review the charge rate for the day calls as total day charge is the most influencial predictor for churn in this model.
- SyriaTel should improve the customer service calls through attentive listening to customers issues, feedbackd and complains and also through offering timely solutions for the same.
- SyriaTel company should reach out to the customers with high numbers of voicemail messages to determine the cause for the voicemail messages surge and know how to adress the same.
- SyriaTel should come up with a tailormade data and voice plan products for the international customers based on their unique needs.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%