SyriaTel Customer Churn analysis

BY : IVET BUYAKI

OVERVIEW

- The project entails building a classifier to assist the client, SyriaTel Telecommunications, in determining whether a customer will soon stop doing business with them. It also focuses on identifying patterns that indicate high probability of customers churning.

OBJECTIVES

The client intends to minimize the resources spent on customers who don't have an intention of staying with the company for the long haul.

The project therefore focuses on:

- Identifying factors that have a high relation with the customer churn rate
- Developing a predictive model that uses above identified factors to predicted possibility of customers churning
- Evaluating the performance of the model against an alternate model and fine tuning it for optimal performance.

DATA UNDERSTANDING

- The data set used is obtained from Kaggle and can be obtained by the following link: https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset/data
- There were no missing values in the data set, it had 3333 rows and 21 columns before any processing was done.

MODELING

- The models used in this case are classification models, as the intention is to determine whether a customer will drop off or not.
- In this particular project we settled on the logistic regression and used the decision tree model as a comparison model.
- The data set had several features and thus after determining their relation with the churn rate, among the influential features were customer service calls, total day charge, total day minutes.

MODELING

A number of steps were undertaken during modeling geared towards having a workable data set fit for creating a well performing model.

These included:

- 1)Data encoding – converting data set to numerical values
- 2)Data scaling – to be on the same scale.
- 3)Data balancing – to prevent bias in prediction due to data set imbalance.

EVALUATION

- Evaluation metrics used was the f1_score as it is a balance point of precision and recall.

1)Precision – depicts how correct the model is in predicting customer churn rate

2)Recall – depicts the accuracy of the model in picking all the customers who actually drop off.

- The logistic regression model was chosen for its advantages such as being less prone to overfitting on data despite the fact that results obtained from the two models were the same.

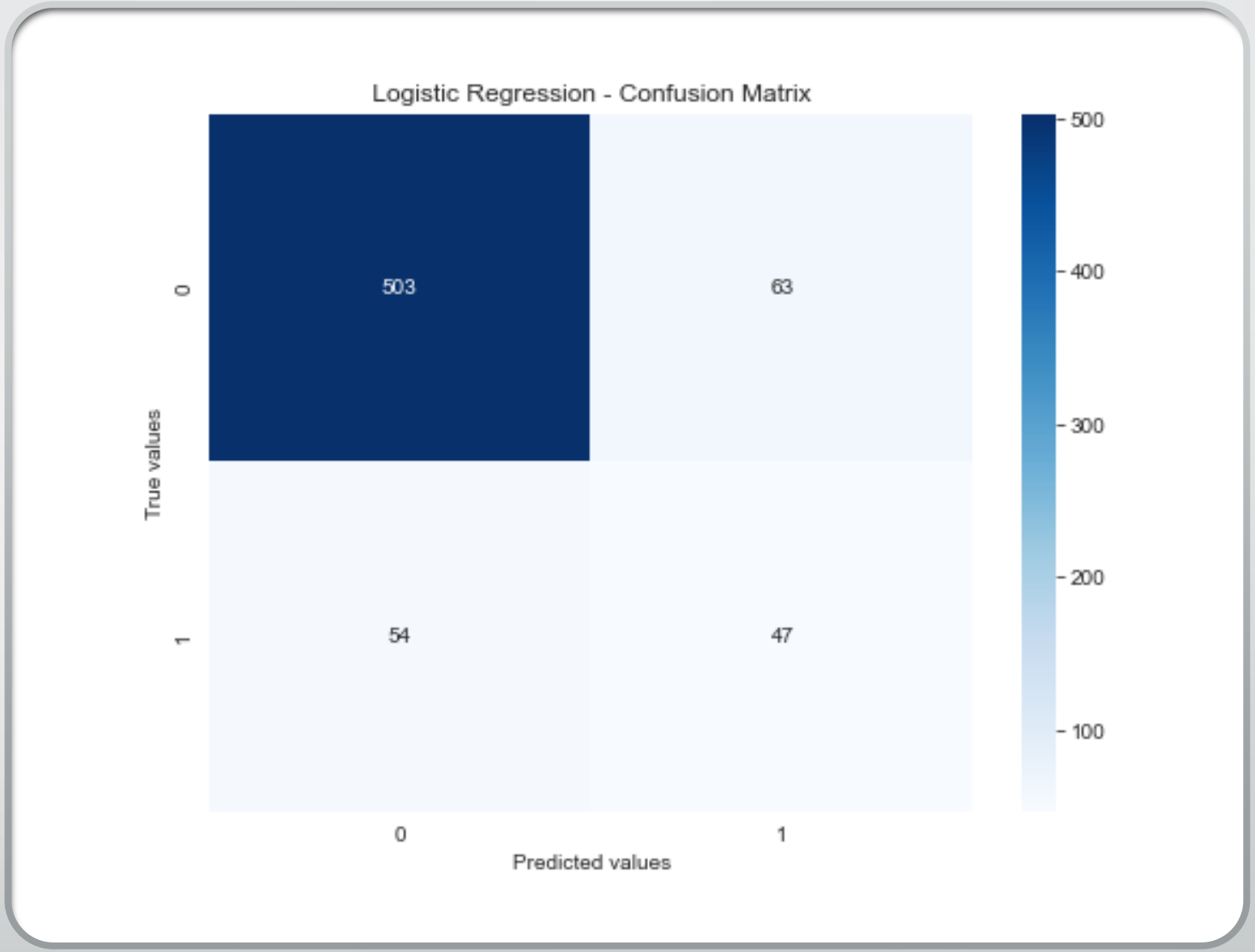
EVALUATION

- The final f1_score obtained from the logistic regression model after all data processing was:

Train data – 0.65

Test data – 0.45

- The accuracy of the model in performing predictions was found to be 0.82.
- The variation in the f1_score can be explained by the unknown nature of the test data.



RECOMMENDATION

- Further tuning be done to improve the performance of the model especially with regards to the f1-score.
- More focus can be put into modifying or upgrading some customer features that have high correlation with the churn rate, such as customer service calls, total day charge, total day minutes
- More classification models can be used to further predict the data set with an aim of attaining the most efficient model.

THANK YOU!