WILLIAM ONSARE MORARA
PHASE 3 PROJECT

SYRIATEL CUSTOMER CHURN ANLYSIS

OVERVIEW

- This project uses machine learning algorithms to build a model that can accurately predict customers who will churn based on the information available in the dataset. The dataset has 20 predictor variables mostly about customer usage patterns. There are 3333 records in this dataset, out of which 483 customers are churners and the remaining 2850 are non-churners. The target variable is 'churn'. Since the target variable is a categorical variable, classification algorithms are used to build the predictive model. Recall is used to evaluate the model's performance.

BUSINESS UNDERSTANDING

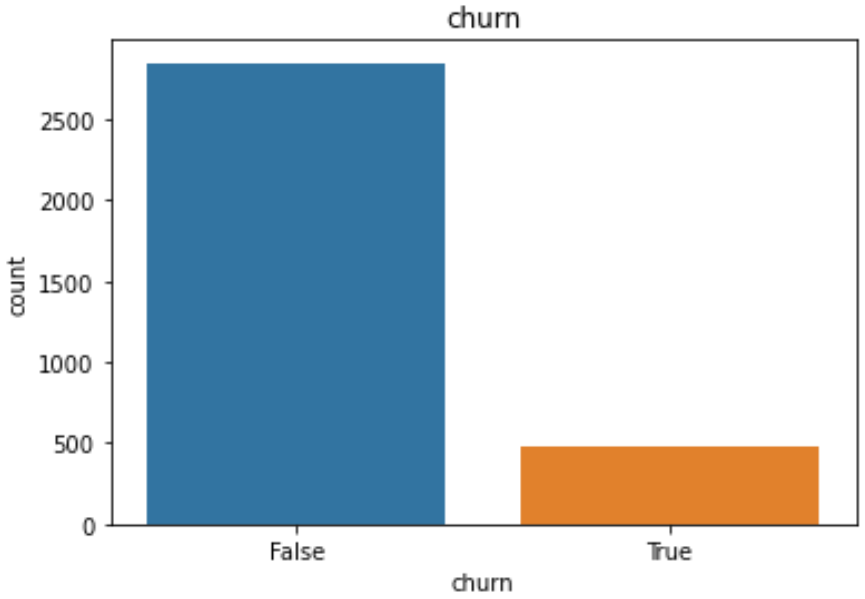
- THE MAIN OBJECTIVE

The main objective of this project is to build and identify machine learning classification model that will correctly classify if a customer will churn or remain in Syriatel

DATA UNDERSTANDING

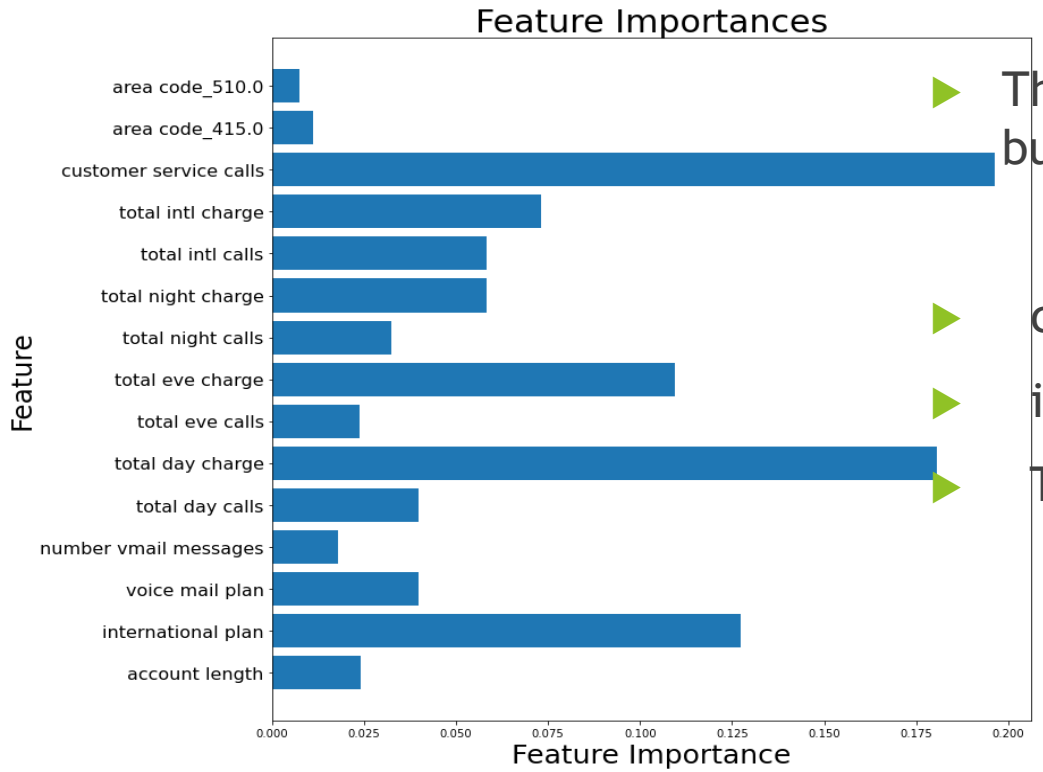
- using the dataset of churning of customers we aim to build the best possible classification model to decrease losses through loss of customers.
- Data is in csv format it contains 3333 unique rows

Customers who churned



- False 2850
- True 483

Important features for our model



- The best features are for model building are:
- customer service calls
- international plan
- Total day charge

Modeling

- In this project we will use classification models because the target variables is categorical.
- Classification models
- logistic regression model(base model)
- k-Nearest-Neighbors
- Decision tree classifier
- random forest

EVALUATION

- Evaluation
- `RECALL` is the time it correctly predicted
- - In the logistic model, the recall for customers who did not churn is 98%, and 18% for customer who churned
- - this model is performing poorly.
- - In the knn model, the recall for customers who did not churn is 99%, and 29% for customer who churned
- - this model is an improvemet of the baseline model but still the recall is low.
- - In the Decision tree model, the recall for customers who did not churn is 99%, and 66% for customer who churned
- - this is a strong model for predicting
- - In the Random forest classifier, the recall for customers who did not churn is 99%, and 66% for customer who churned
- - this model is strong

Tuned model evaluation

- `RECALL` is the time it correctly predicted
- - In the logistic model, the recall for customers who did not churn is 78%, and 70% for customer who churned
- - this model is performing better than the baseline model.
- - In the knn model, the recall for customers who did not churn is 84%, and 65% for customer who churned
- - this model is an improvemet of the baseline model but still the recall is low.
- - In the Decision tree model, the recall for customers who did not churn is 92%, and 80% for customer who churned
- - this is a very strong model for predicting
- - In the Random forest classifier, the recall for customers who did not churn is 96%, and 75% for customer who churned
- - this model is strong

BEST model

- The best model for this project is decision tree

Recomendations

- we recommend use of decision tree model because it had the best recall
- we recommend use of this feature
- using more balanced data
-
- look into such features such as voicemail plan and international plan because they have high churning.

NEXT STEPS

- Deploy the model to predict

THANK YOU