SyriaTel Classification Analysis

By Susanna Han

DATA COLLECTED:

- Total **intl** minutes
- Total **intl** calls
- Total **intl** charge

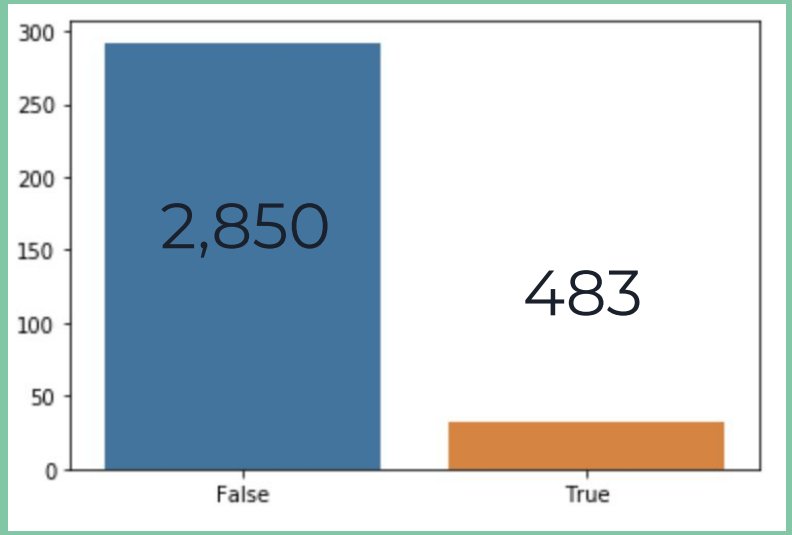
- State
- Account length
- Area code
- Phone number
- International plan
- Voicemail plan
- Number vmail messages
- Customer service calls
- **CHURN**

- 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

3,333 total customers



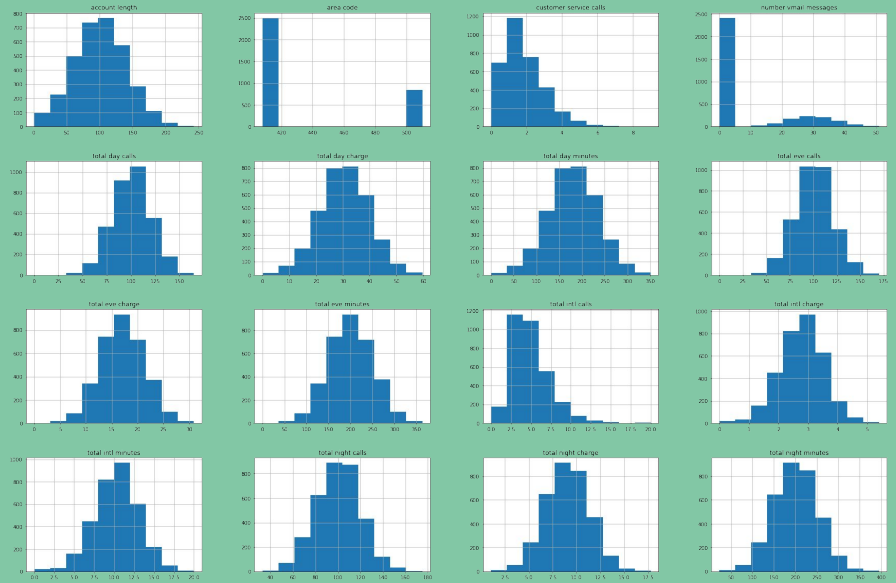
Churn

14.5%

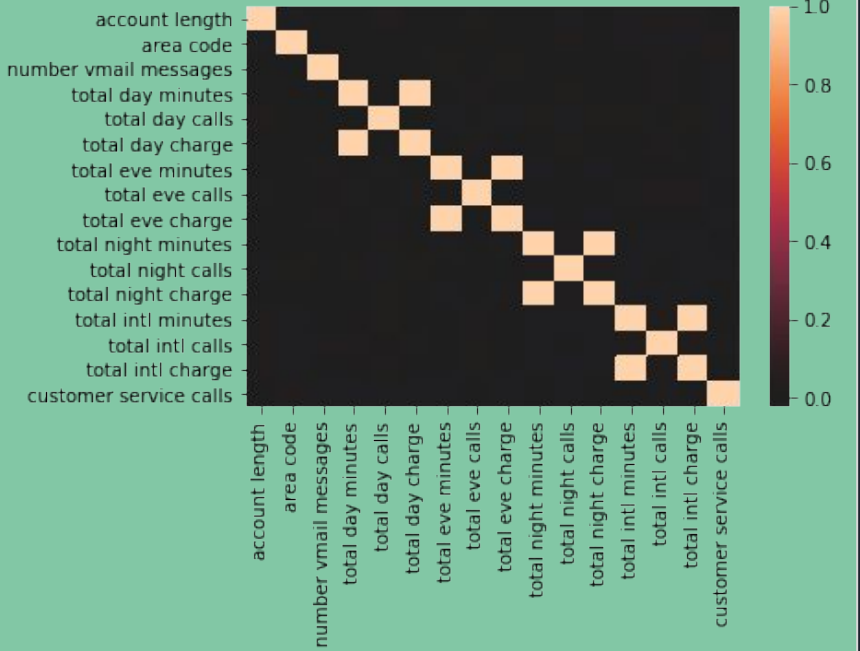
of customers no longer do business with SyriaTel

$$\frac{\text{Total \# of customer churn}}{\text{Total \# of customers}}$$

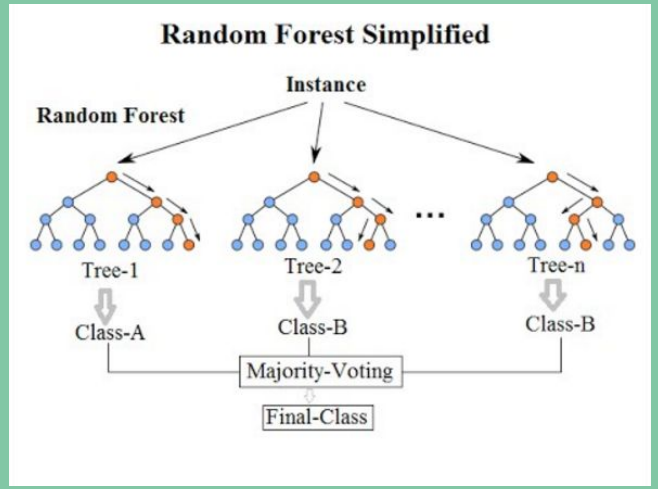
Normal distribution



Correlation



Methods :



Random Forest is an algorithm that consist of many decision trees using random subsets of features that averages the outcome to make predictions.

RandomizedSearchCV is an algorithm that takes the input of parameters and attempts every combination and outputs the best hyper parameters for the model.

Random Forest - base model

test score : **0.936**

N_estimators: 100
- # of trees

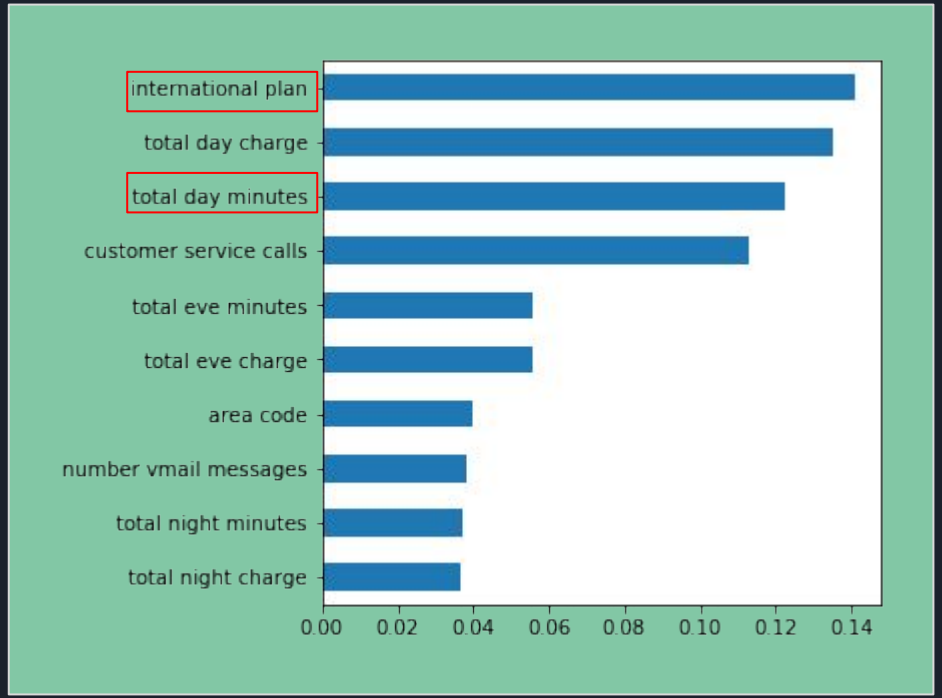
Min_samples_split: 2
- Min required to be a internal node.

Min_samples_leaf: 1
- Min required to be a leaf node.

Max_features: auto
- # features to consider

Max_depth: None

Criterion: gini
- calculates impurity



Random Forest - tuned model

test score : **0.952**

N_estimators: 1000
- # of trees

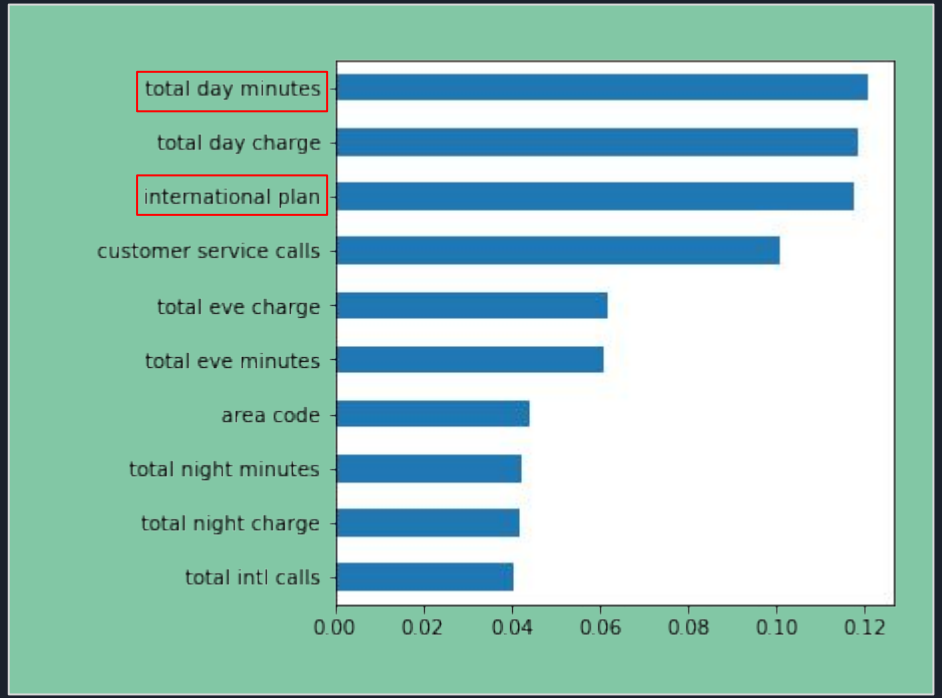
Min_samples_split: 2
- Min required to be a internal node.

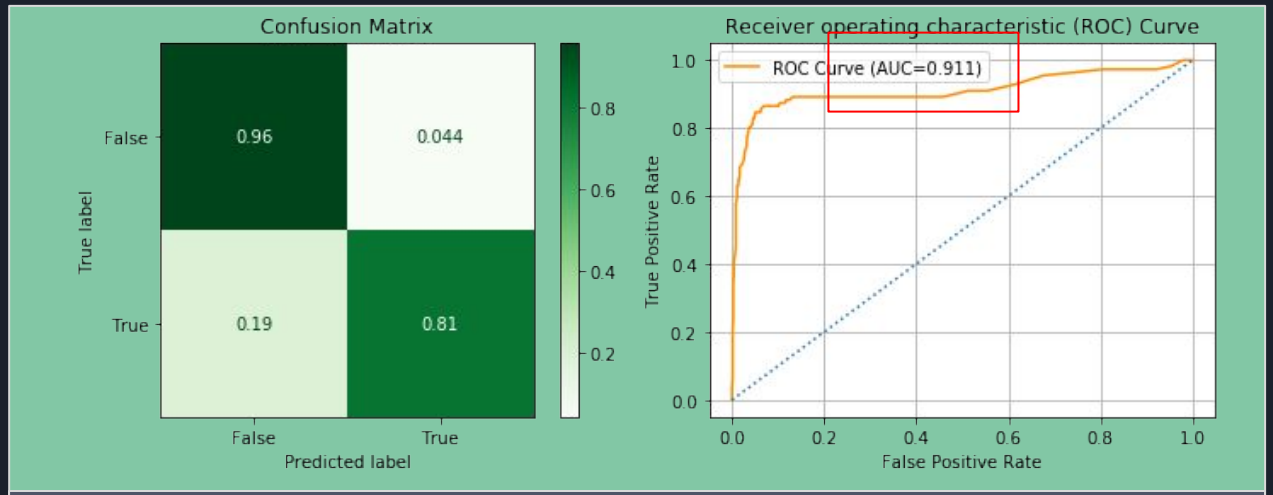
Min_samples_leaf: 1
- Min required to be a leaf node.

Max_features: auto
- # features to consider

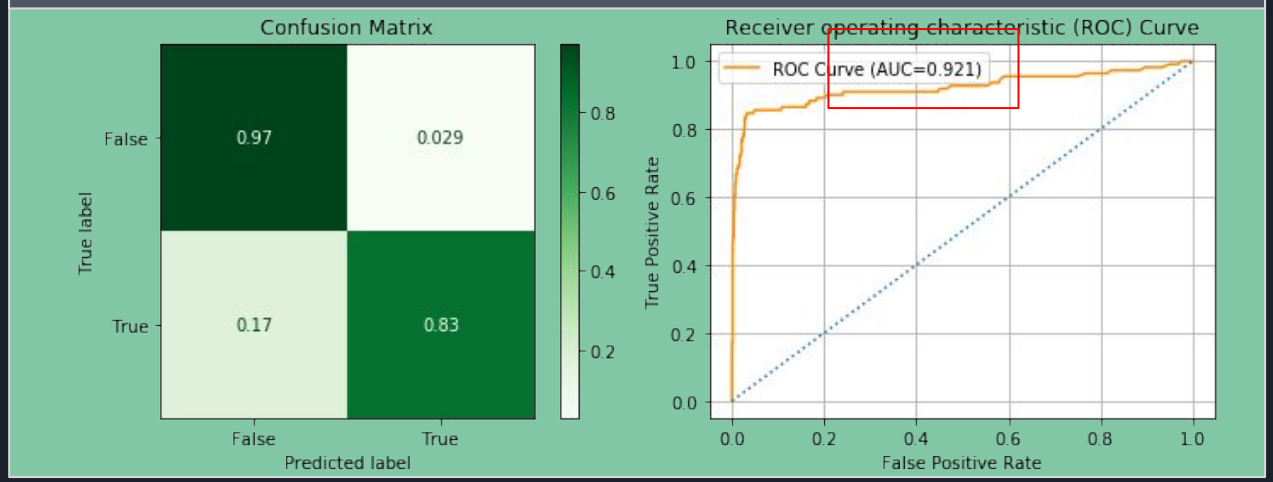
Max_depth: None

Criterion: entropy
- calculates information gain





Base Model



Tuned Model

Insight and Recommendations

1. Customers who use more minutes during the day and get charged are more likely to stop doing business.
    a. lowering total day charge.

2. Customers that have a voice mail set up are more likely to continue business.
    a. Assist customers with setting up a voicemail.

3. Customers who have an international plan are more likely to discontinue doing business.

Future Work:

- Fit the data to GradiantBoosting and compare different models.
- Create more vibrant visuals to help analyze data.
- Compare GridSearchCV with RandomizedSearchCV

Thank you!