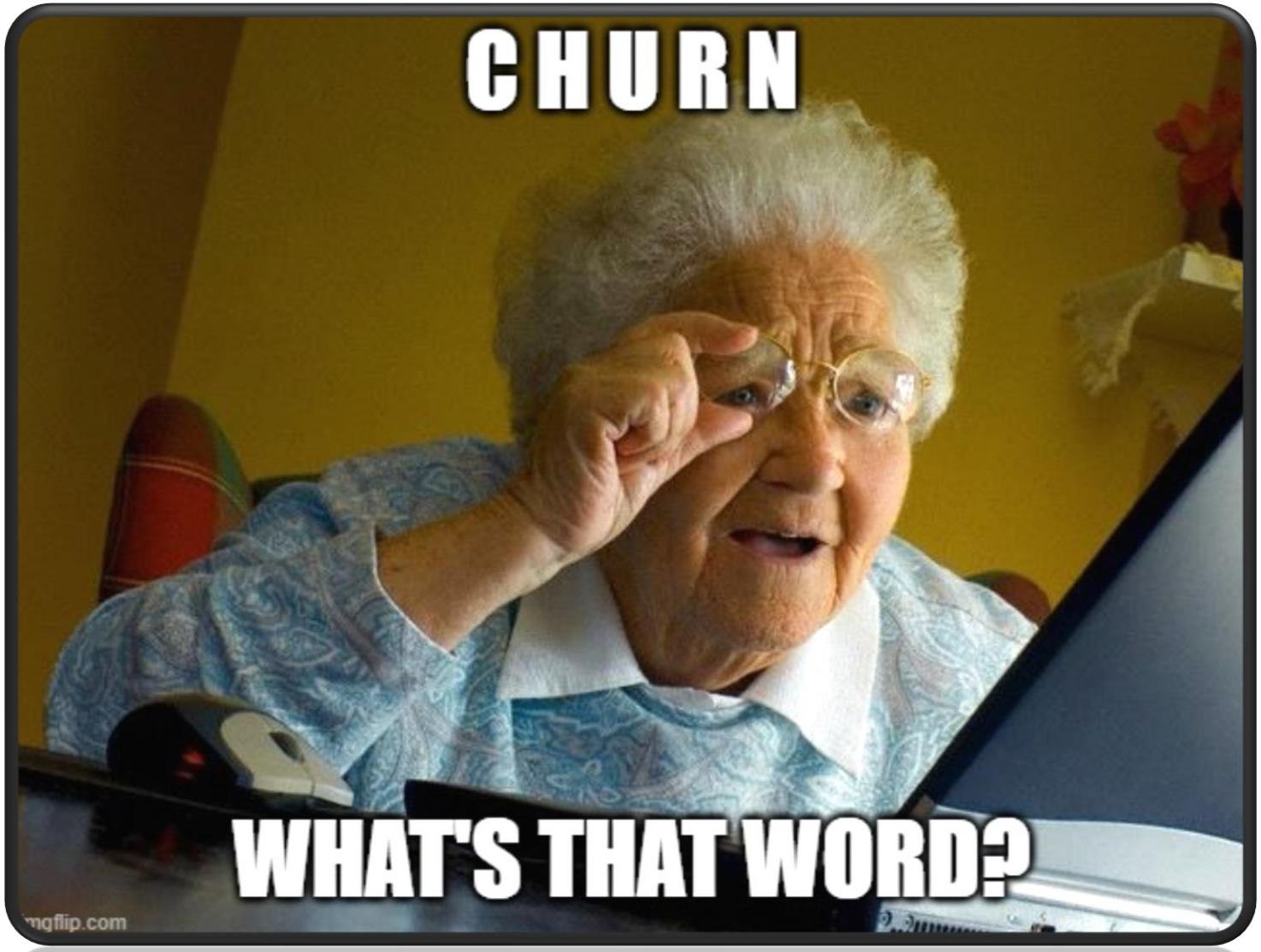
THINKFUL Data Science

Capstone Project 2

Prediction of Customer Churn Rates for a Telecommunications Business.

Henry Ssembatya

Introduction



CUSTOMER CHURN

Percentage of customers that stopped using your company's product or service during a certain time frame.

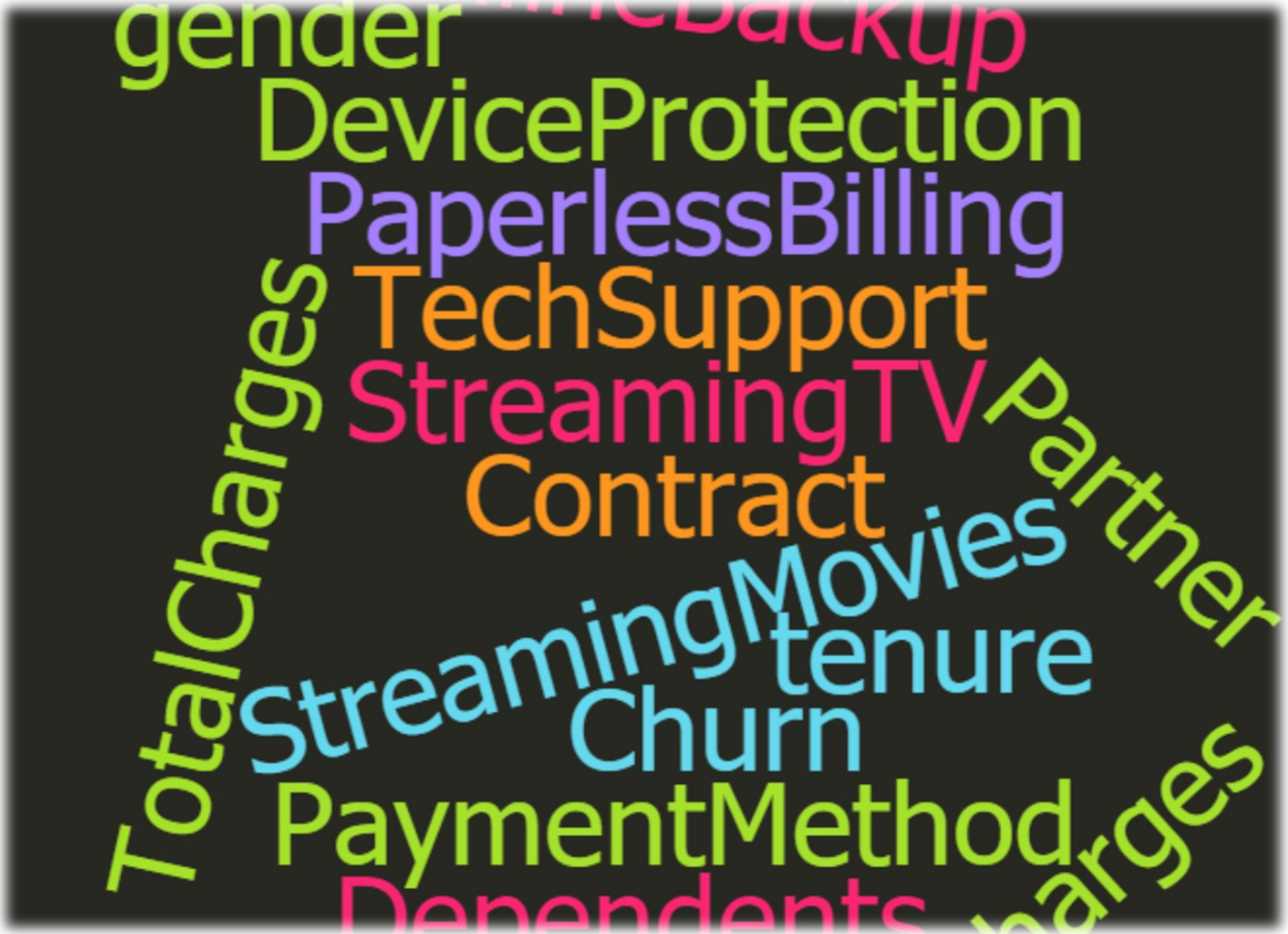
Source:@swethamaresan

Why its important!

- Costs more to acquire new customers than it does to retain.
- More customer retention, more profit, more growth.
- Indicates whether your product, service, and team is succeeding.

Dataset

Features



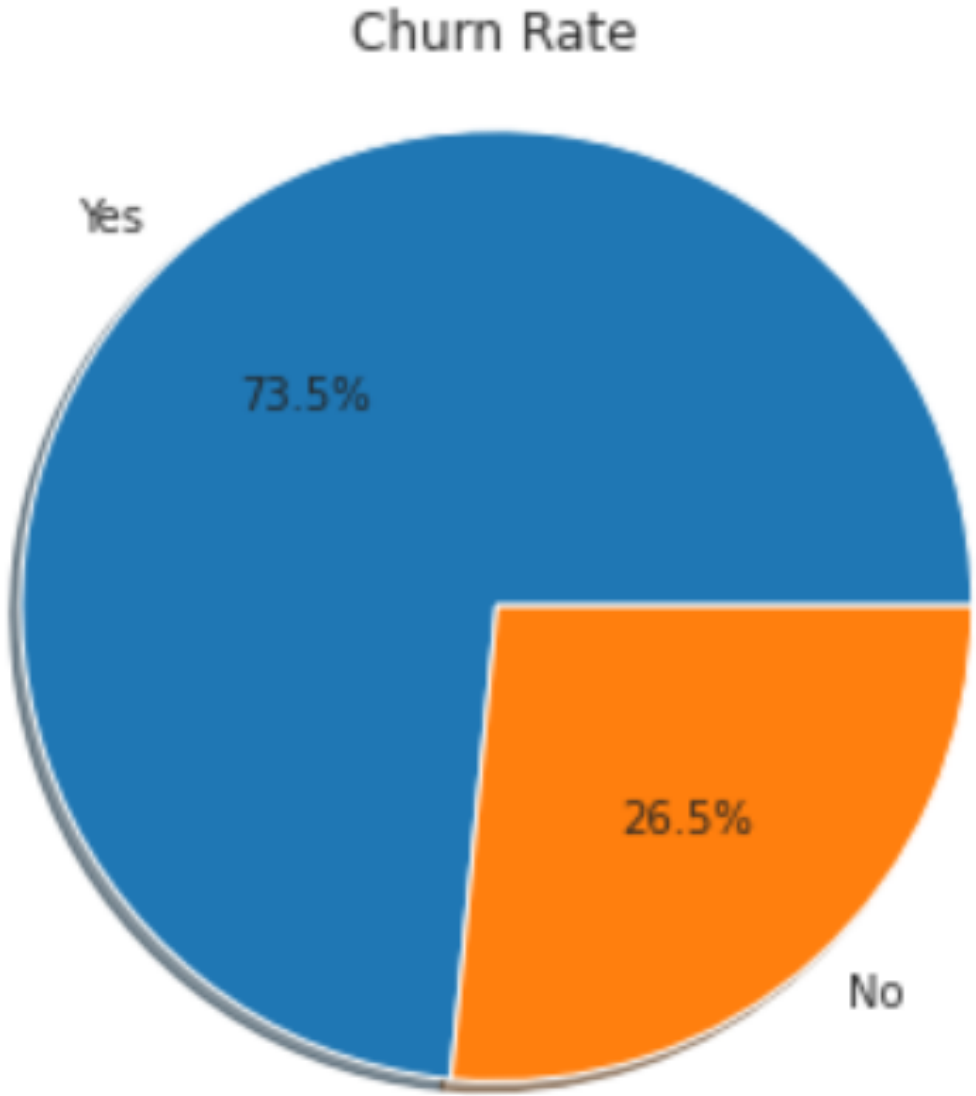
Size:

7043 Instances

x21 Features

SOURCE: Kaggle

Target Variable: Churn Rate



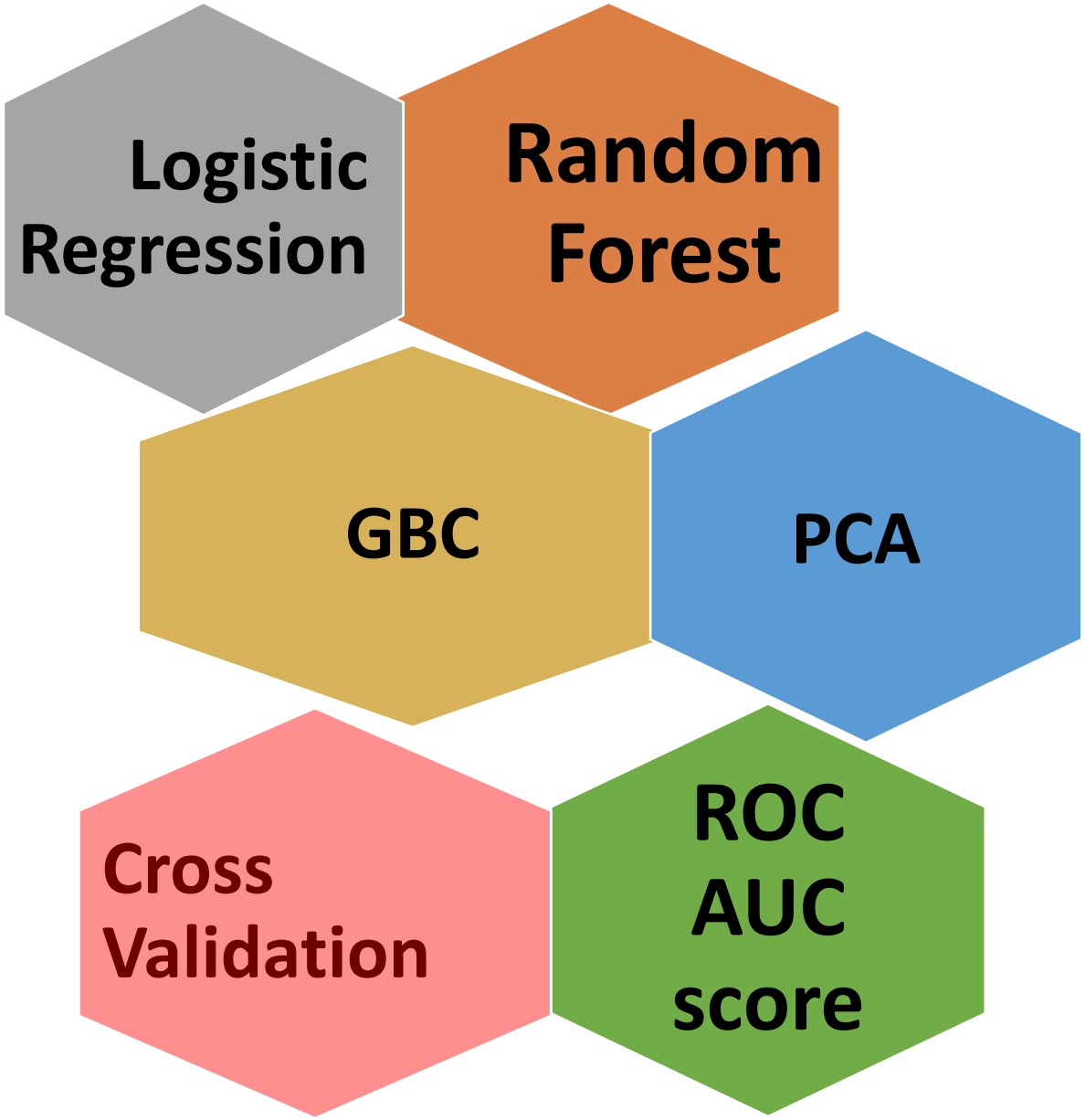
Observation:

- Imbalanced

Meaning:

- Use metrics like AUC instead of accuracy.
- Resampling

Machine Learning Toolkit



CLASSIFICATION PROBLEM

Analysis

Logistic Regression

Results

- Optimization Parameters

```
param_distributions = random_grid
n_iterations = 100
Cross Validation = 5  Fold
verbose=2
```

ROC-AUC
Score: 0.844

0.2% Improvement

ROC-AUC
Score: 0.846

Analysis

Random Forest

Results

- Optimization Parameters

```
param_distributions = random_grid
n_iterations = 100
max_depth=9
max_features="auto"
```

ROC-AUC
Score: 0.794

6% Improvement

ROC-AUC
Score: 0.845

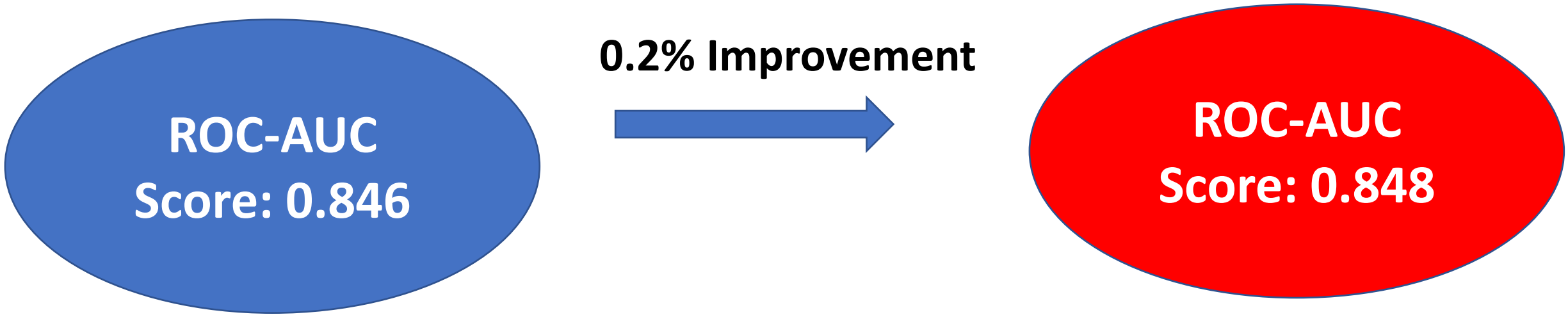
Analysis

Gradient Boost Classifier

Results

- Optimization Parameters

```
param_distributions = random_grid
n_iterations = 100
max_depth=3
max_features=8
```



Analysis

Comparison of 3 models

0.846 Logistic Regression

•

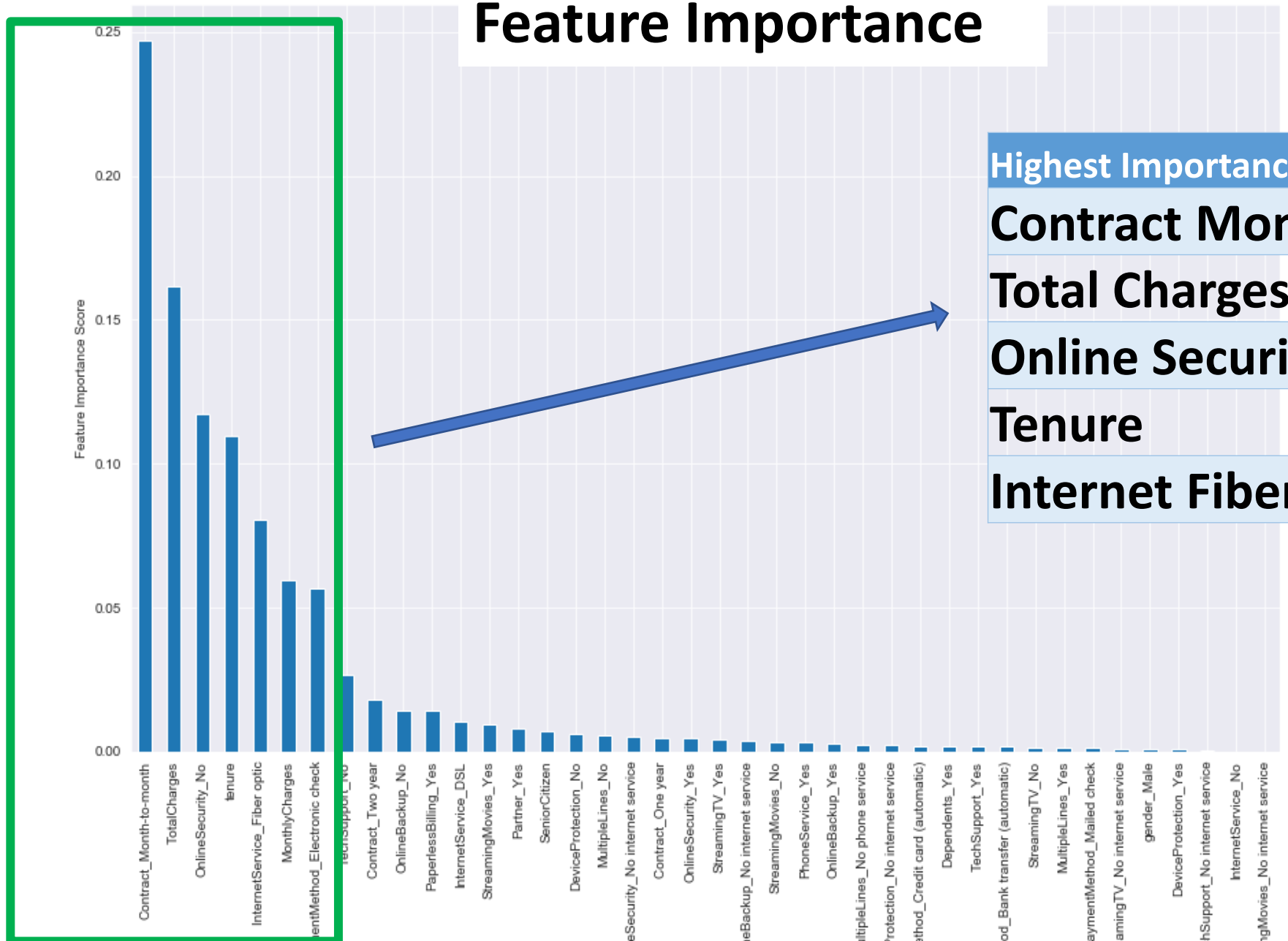
0.845 Random Forest

•

0.848 GBC

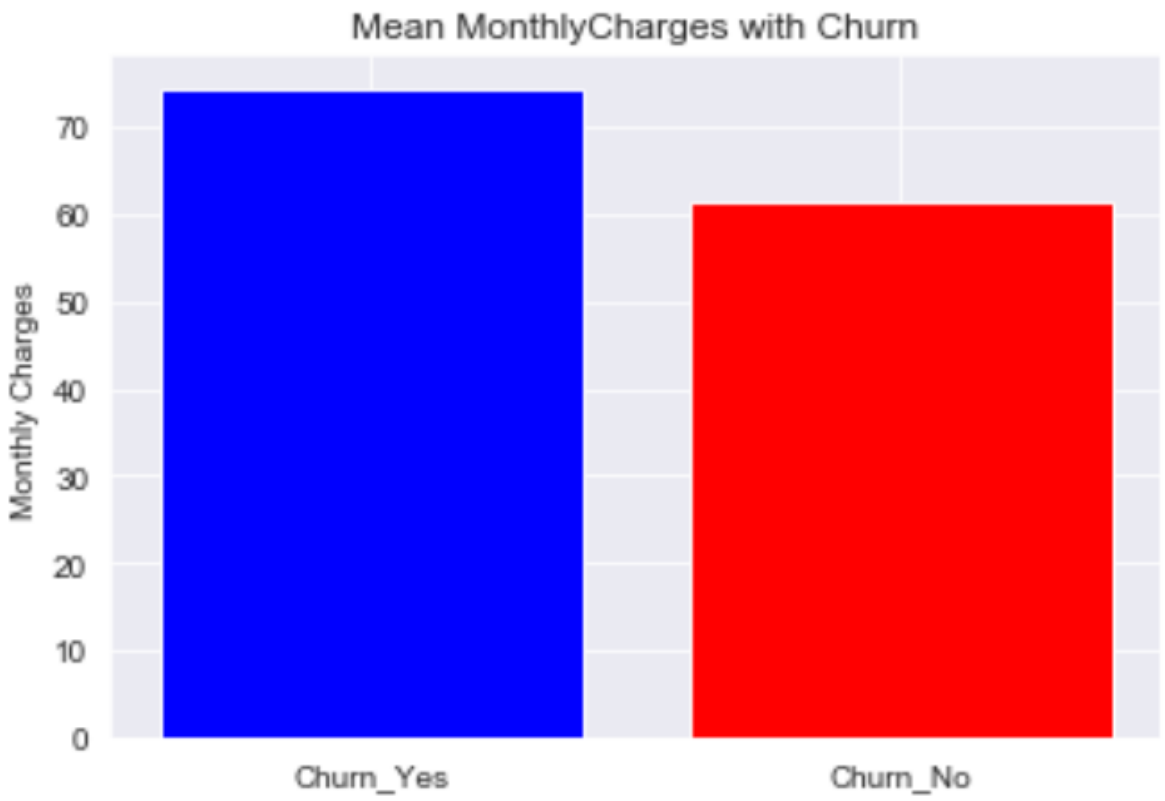
Results

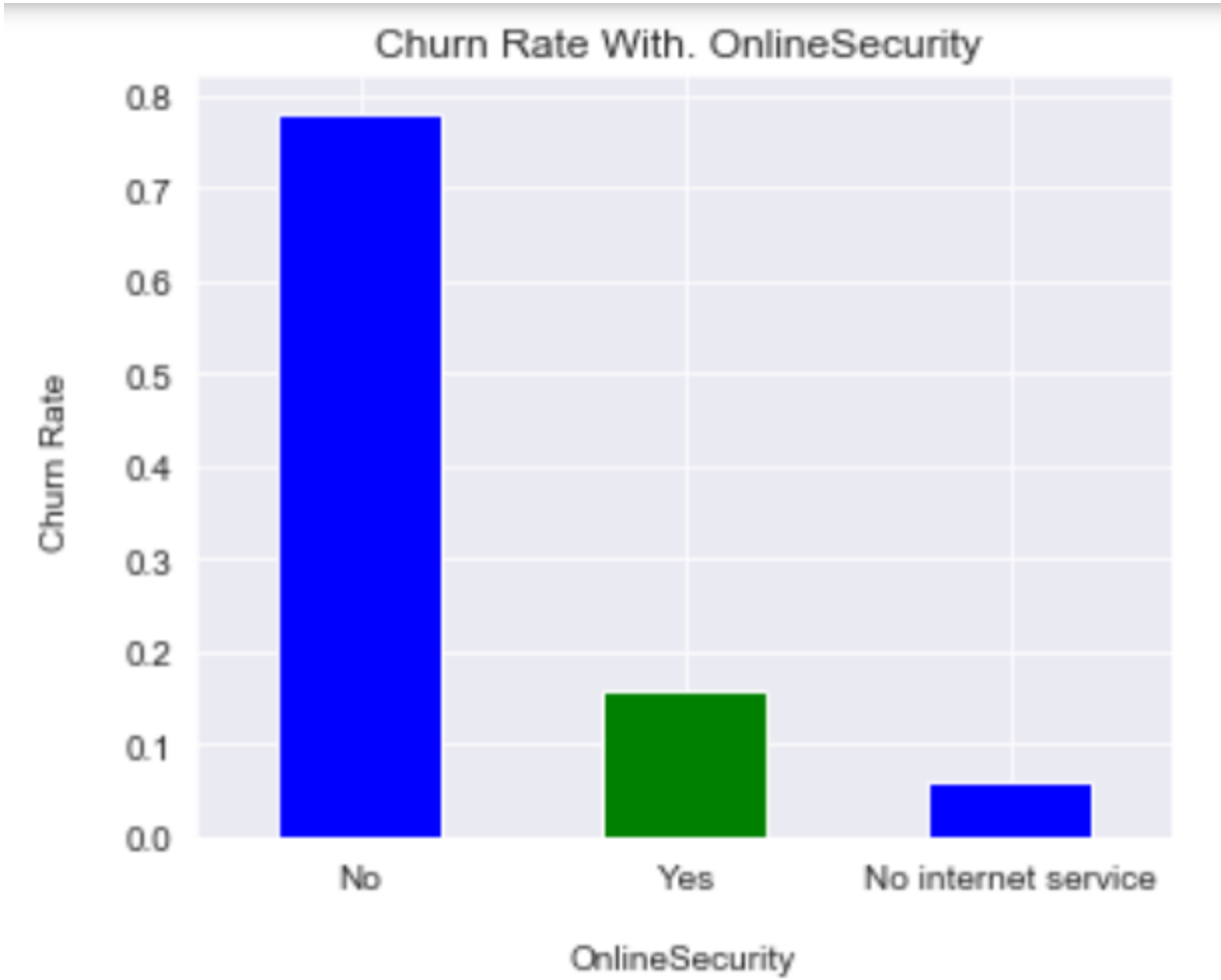
Feature Importance



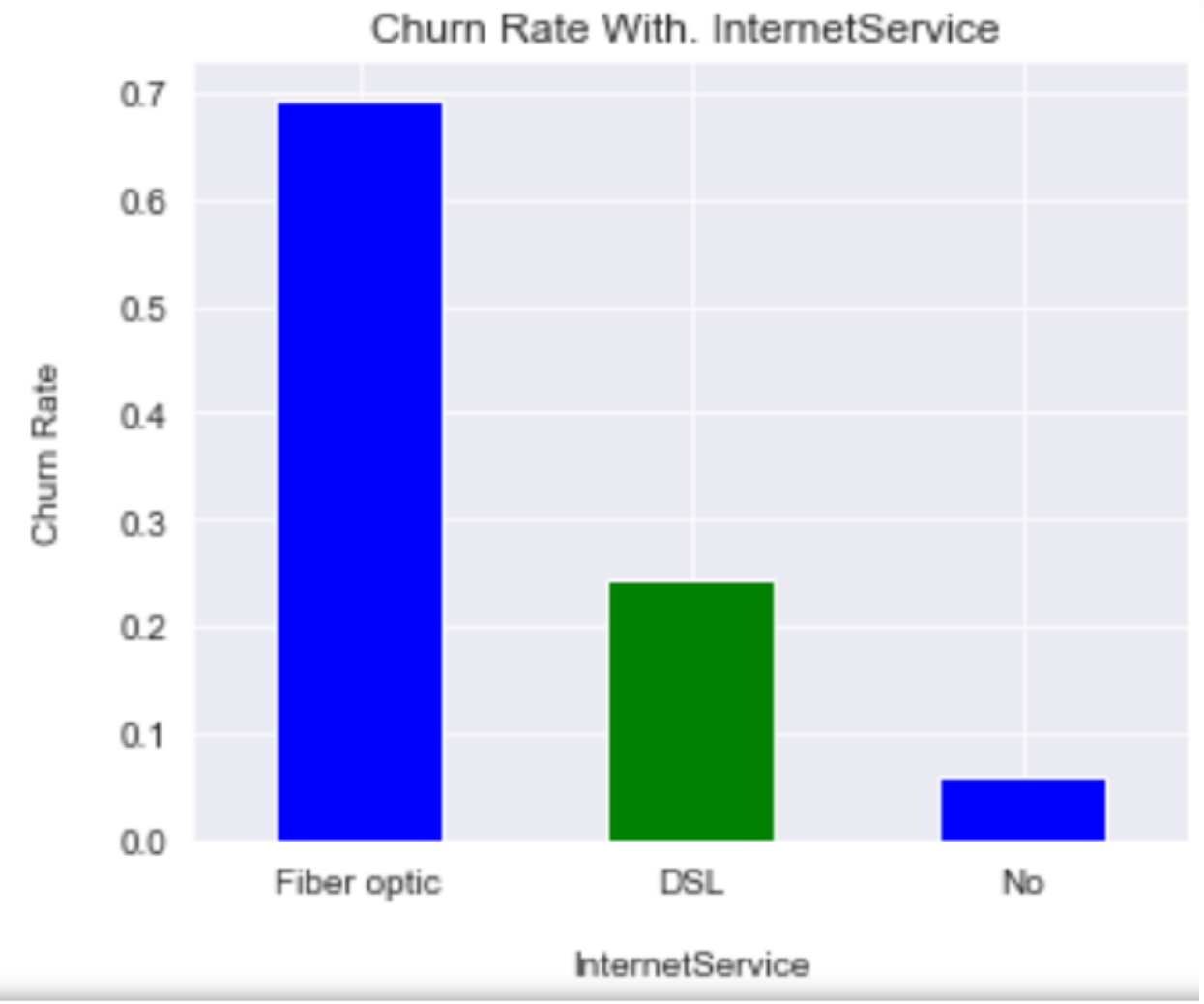
Highest Importance for Model
Contract Month to Month
Total Charges
Online Security
Tenure
Internet Fiber Optic

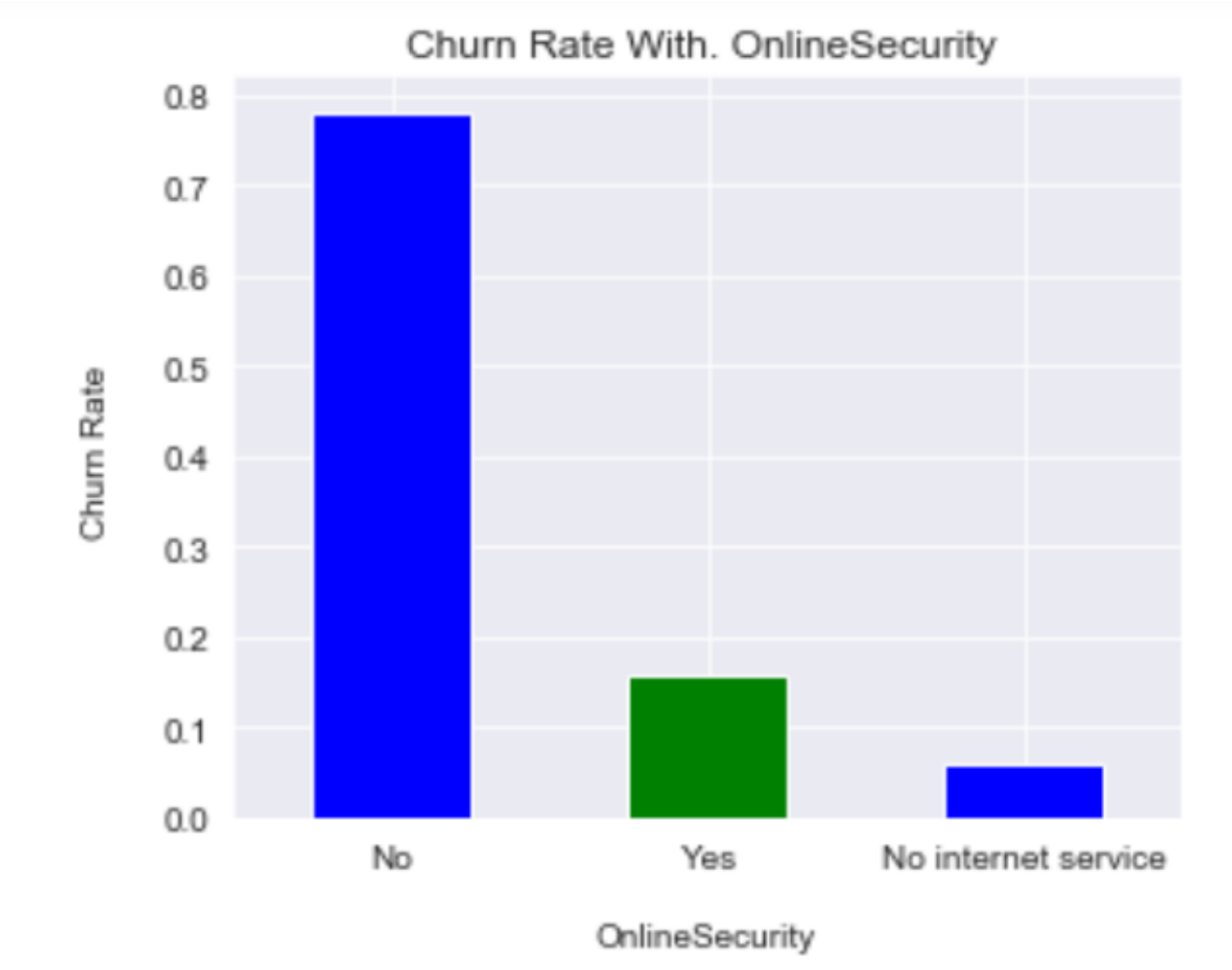
Analysis


Mean MonthlyCharges with Churn
Monthly Charges
0
10
20
30
40
50
60
70
Churn_Yes
Churn_No


Churn Rate With. OnlineSecurity
Churn Rate
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
No
Yes
No internet service
OnlineSecurity

Analysis


Churn Rate With. InternetService
Churn Rate
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Fiber optic
DSL
No
InternetService


Churn Rate With. OnlineSecurity
Churn Rate
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
No
Yes
No internet service
OnlineSecurity

Interpretation of Coefficients

Statistical Analysis: Odds Ratio

const	1.173221
Contract_Month-to-month	1.147518
TotalCharges	0.999983
OnlineSecurity_No	1.123607
InternetService_Fiber optic	1.250110
tenure	0.997328

Conclusions

- Gradient Boost Classifier method is most effective in predicting the churn.
- Hyperparameter tuning effective in improving model performance.
- Feature analysis, we have effectively identified the features that have greater impact

Recommendations

- Annual contracts Vs Monthly Contracts?
- Online security for customers.
- Review monthly charges.
- Internet service: Fiber Optic VS DSL?
- More data on CHURN=NO

Future Work

- More powerful tools like Ensemble Methods.
- More research into the company to identify
- other pertinent causes of Churn.
- Understanding of terms like “Tenure”

Thanks

谢谢

Gracias

Henry Ssembatya