Predicting Gold Glove Award Winners with Adaptive Boosting Machine Learning

Lucas Kelly

Problem Statement

Errors: 0
Put outs: 271
Fielding %: 100

Errors: 5
Put outs: 299
Fielding %: 98.4

Errors: 7
Put outs: 355
Fielding %: 98.1

The Value of Machine Learning

- Utilize advanced metrics
- Take into consideration adjusted metrics
- Improving over time

Methodology

- Data (web scraping, combined advanced/traditional statistics)
- Model selection (hyperparameter tuning, PCA, Gridsearch, multicollinearity removal, feature selection)
- Unsupervised Learning: Adaptive Boosting Ensemble Method

Findings

Adaptive Boosting:

Model Accuracy - 89%

Precision - 58%

Recall - 29%

F1 Score - 39%

Findings

GG Winner!

Finalist

GG WInner!

Finalist

GG Winner!

GG Winner!

Future Work

- Investigate weight by position
- Finer tuning
- Experiment with XGBoost Algorithms
- Test model on upcoming 2019 awards

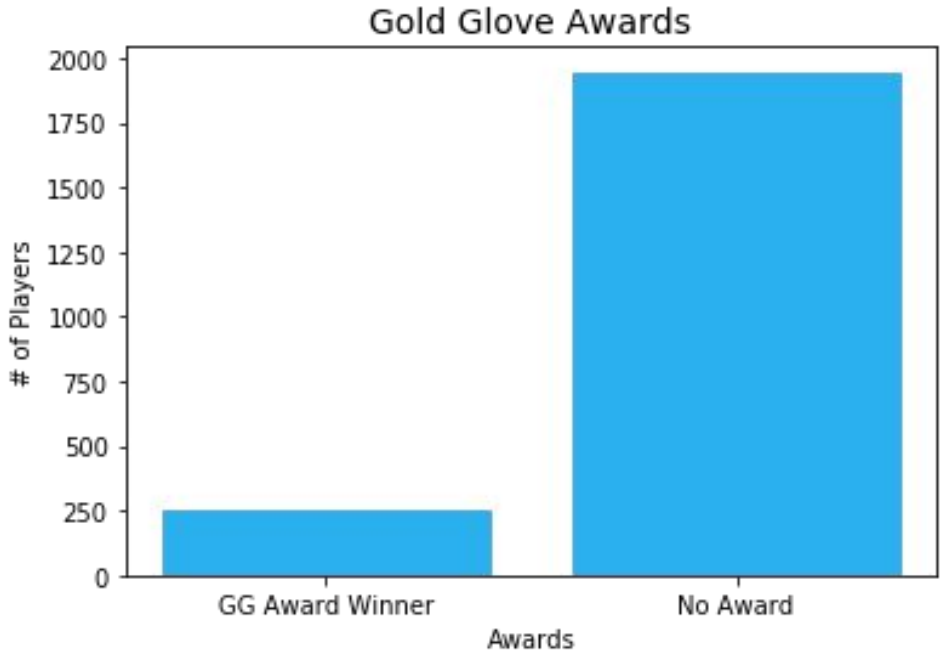
Thank you!

www.lucaskellydataportfolio.weebly.com

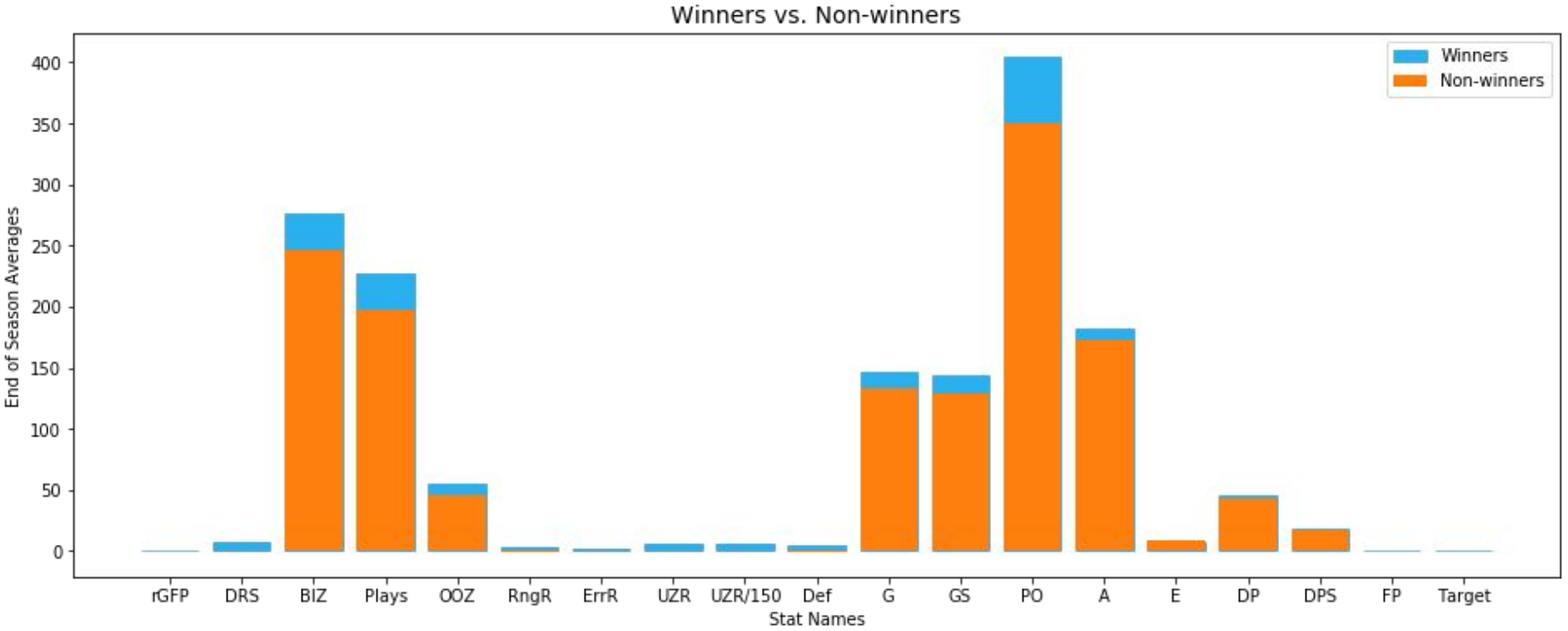
www.github.com/lucaskelly49

https://www.linkedin.com/in/lucaskelly49/

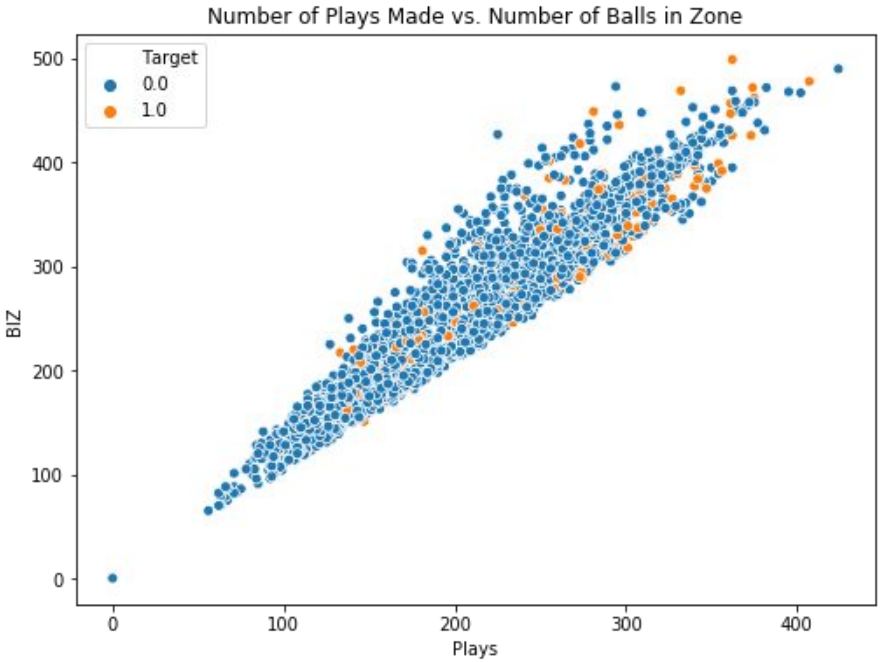
EDA: GG Awards vs. Non-awards

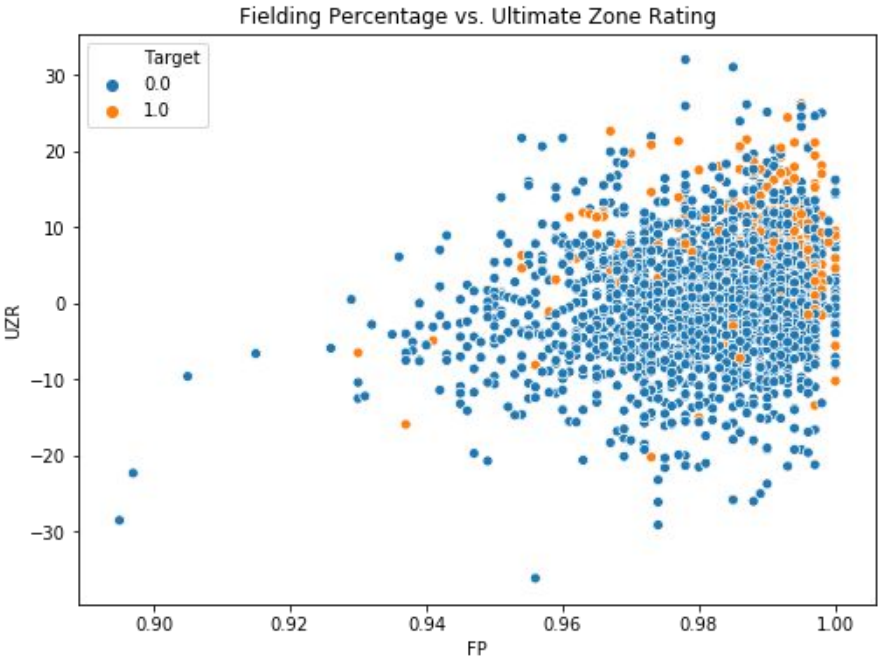


EDA: GG Awards vs. Non-awards

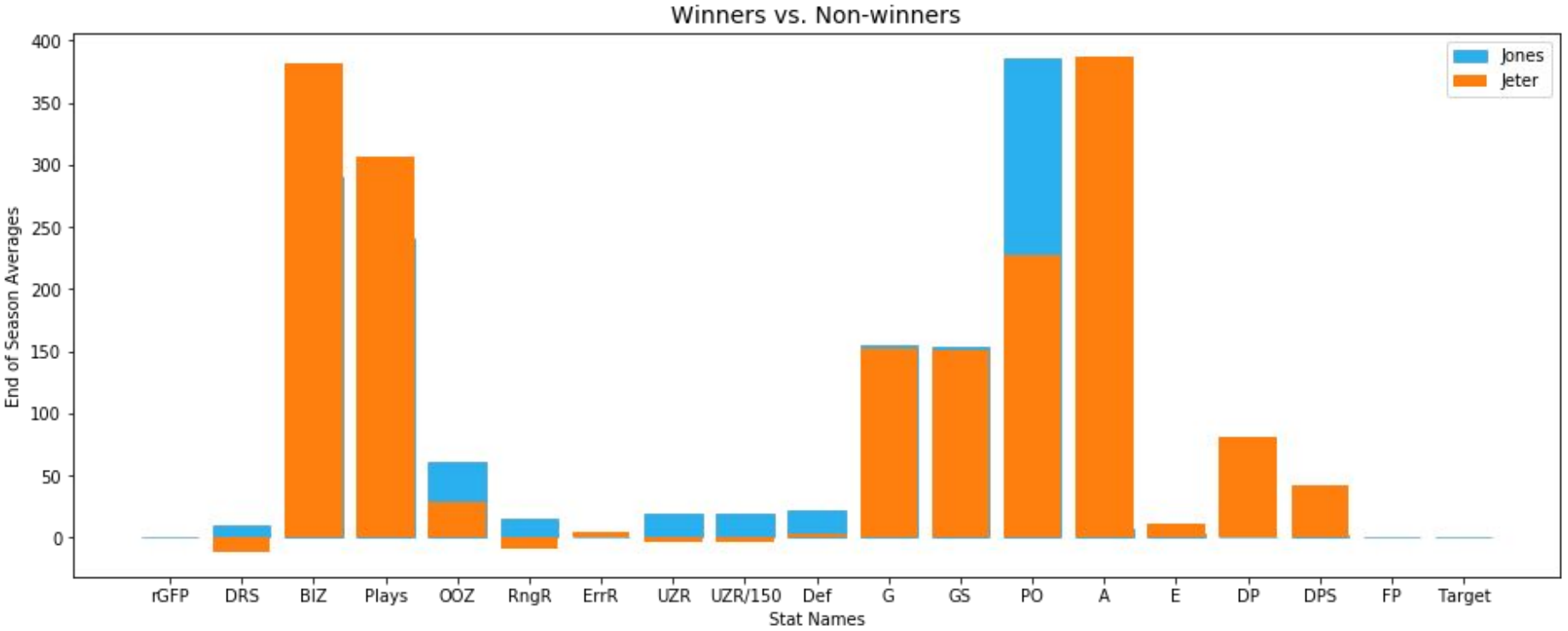


EDA: Stat correlations





EDA: Influence of position



KNN evaluation metrics

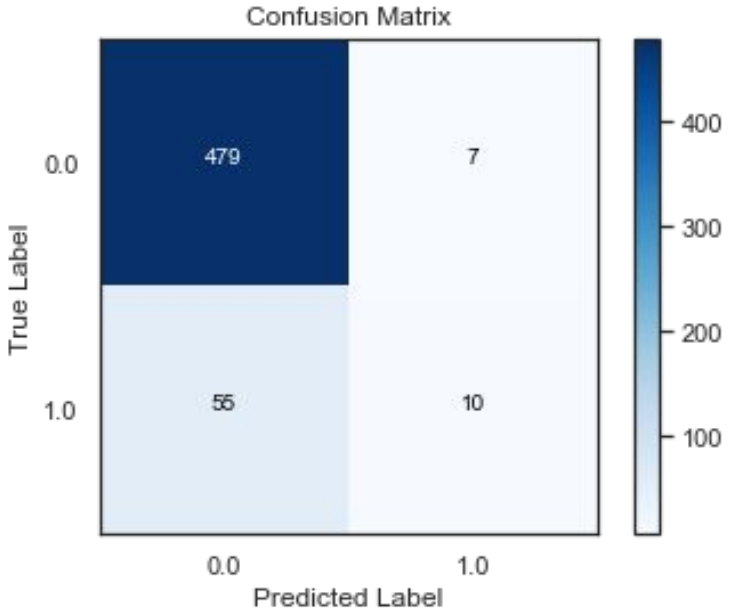
```
Classification Report K-Nearest Neighbors:
              precision    recall  f1-score   support

         0.0       0.90      0.99      0.94       486
         1.0       0.59      0.15      0.24        65

   micro avg       0.89      0.89      0.89       551
   macro avg       0.74      0.57      0.59       551
weighted avg       0.86      0.89      0.86       551

----------------------------------------------------------

KNN
----------------------------------------------------------
Training Accuracy: 89.76%
Test Accuracy:     88.75%
```



Decision Tree evaluation metrics

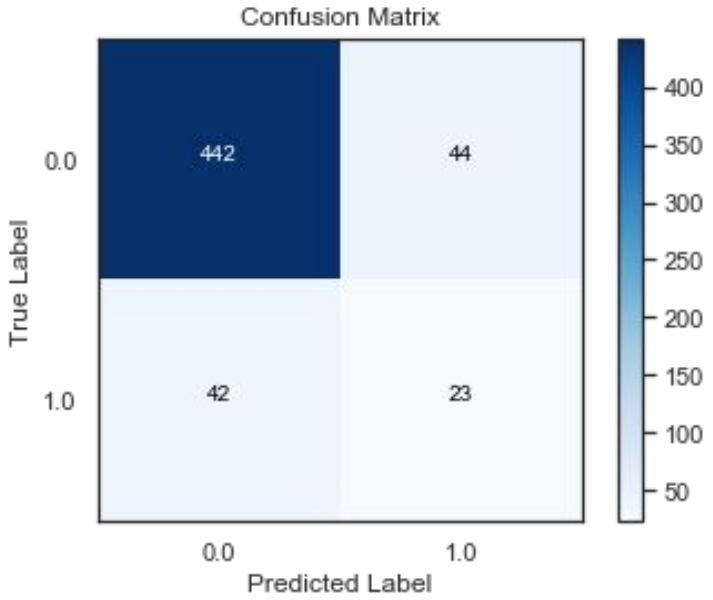
```
Classification Report Vanilla Decision Tree:
              precision    recall  f1-score   support

         0.0       0.91      0.91      0.91       486
         1.0       0.34      0.35      0.35        65

   micro avg       0.84      0.84      0.84       551
   macro avg       0.63      0.63      0.63       551
weighted avg       0.85      0.84      0.84       551

----------------------------------------------------------

Vanilla Decision Tree
----------------------------------------------------------
Training Accuracy: 100.0%
Test Accuracy:     84.39%
```



Random Forest evaluation metrics

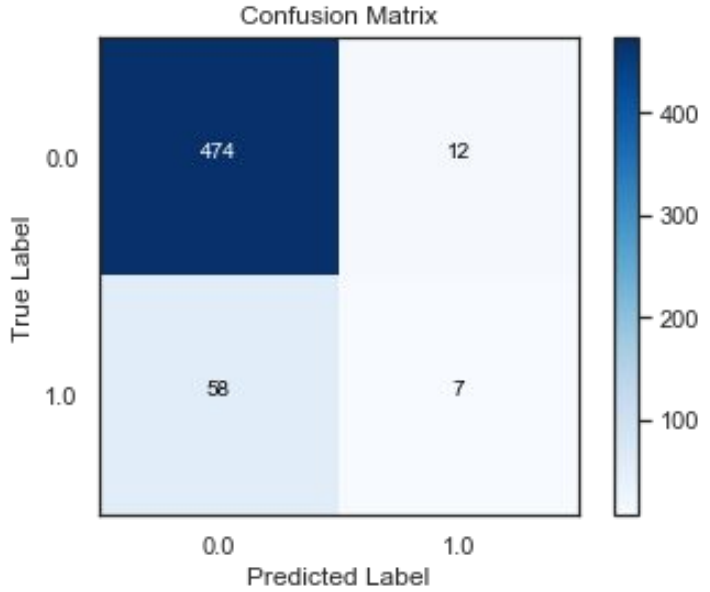
```
Classification Report Random Forest:
              precision    recall  f1-score   support

         0.0       0.89      0.98      0.93       486
         1.0       0.37      0.11      0.17        65

   micro avg       0.87      0.87      0.87       551
   macro avg       0.63      0.54      0.55       551
weighted avg       0.83      0.87      0.84       551

------------------------------------------------------

Random Forest
------------------------------------------------------
Training Accuracy: 99.03%
Test Accuracy:     87.3%
```



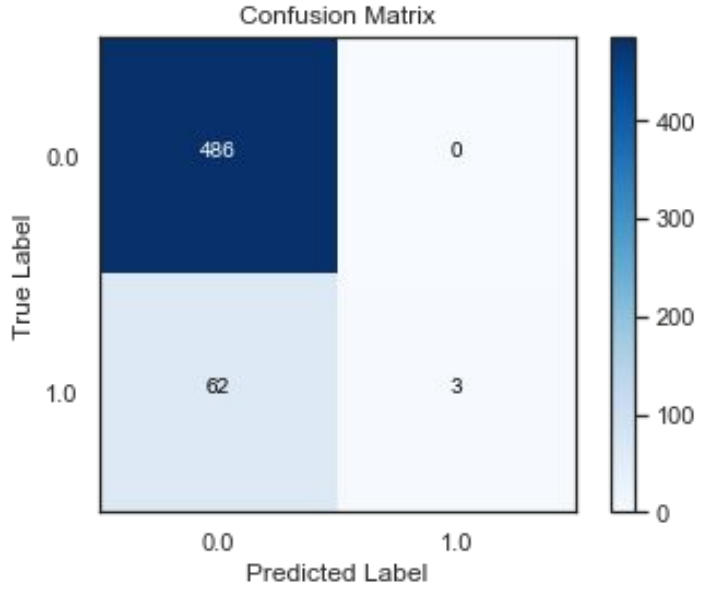
Tuned Random Forest w/ Grid Search evaluation metrics

Classification Report Tuned Random Forest:

	precision	recall	f1-score	support
0.0	0.89	1.00	0.94	486
1.0	1.00	0.05	0.09	65
micro avg	0.89	0.89	0.89	551
macro avg	0.94	0.52	0.51	551
weighted avg	0.90	0.89	0.84	551

Tuned Random Forest

Training Accuracy: 89.22%
Test Accuracy: 88.75%

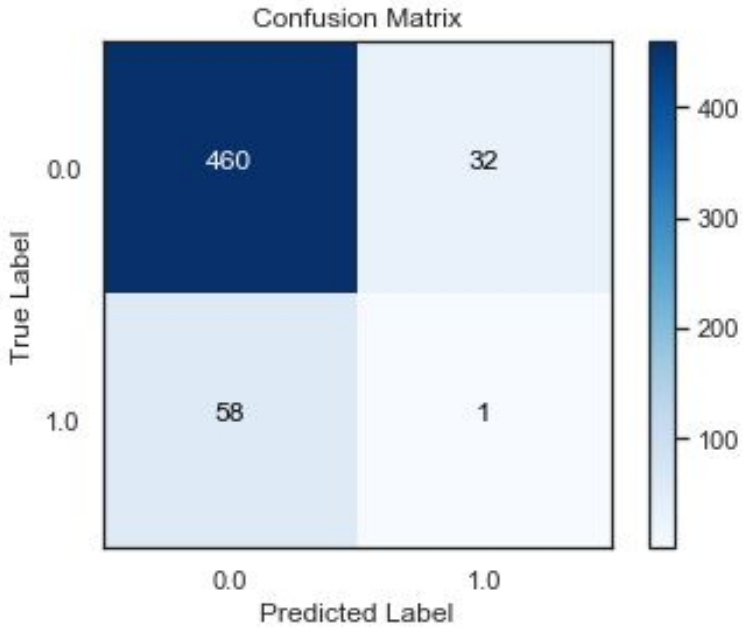


Adaboost evaluation metrics

	precision	recall	f1-score	support
0.0	0.91	0.97	0.94	486
1.0	0.58	0.29	0.39	65
micro avg	0.89	0.89	0.89	551
macro avg	0.74	0.63	0.66	551
weighted avg	0.87	0.89	0.88	551

Adaboost

Training Accuracy: 89.76%
Test Accuracy: 89.11%



Gradient Boost evaluation metrics

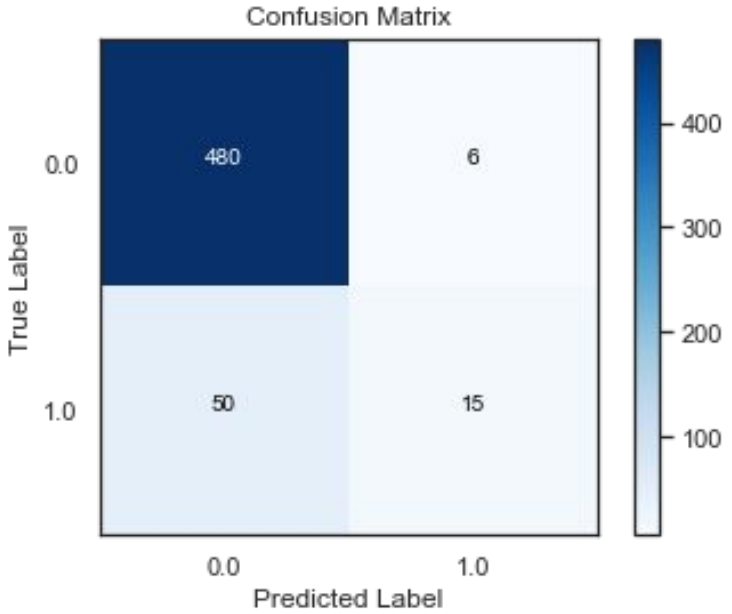
```
Classification Report Gradient Boost:
              precision    recall  f1-score   support

         0.0       0.91      0.99      0.94       486
         1.0       0.71      0.23      0.35        65

   micro avg       0.90      0.90      0.90       551
   macro avg       0.81      0.61      0.65       551
weighted avg       0.88      0.90      0.87       551

----------------------------------------------------

Tuned Random Forest
----------------------------------------------------
Training Accuracy: 94.37%
Test Accuracy:     89.84%
```



Model Comparisons

	Model	Training ACC	Testing ACC	Testing Precision	Testing F1
0	KNN	0.899455	0.892922	0.666667	0.289157
0	Vanilla Decision Tree	1.000000	0.836661	0.307692	0.307692
0	Random Forest	0.978195	0.889292	0.642857	0.227848
0	Vanilla Decision Tree	1.000000	0.836661	0.307692	0.307692
0	Random Forest	0.981829	0.874773	0.300000	0.080000
0	Tuned Random Forest	0.890975	0.882033	0.000000	0.000000
0	Adaboost	0.897638	0.891107	0.575758	0.387755
0	Gradient Boost	0.942459	0.894737	0.705882	0.292683
